

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
DEPARTAMENTO DE ESTATÍSTICA

Francisco Felipe de Queiroz

**Estudo sobre má especificação na família de  
posição e escala**

Natal - RN

25 de Novembro de 2016

Francisco Felipe de Queiroz

## **Estudo sobre má especificação na família de posição e escala**

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Dione Maria Valença

Natal - RN

25 de Novembro de 2016

Catálogo da Publicação na Fonte. UFRN / SISBI / Biblioteca Setorial  
Centro de Ciências Exatas e da Terra – CCET.

Queiroz, Francisco Felipe de.

Estudo sobre má especificação na família de posição e escala / Francisco Felipe de Queiroz. - Natal, 2016.

56f. : il.

Orientadora: Profa. Dra. Dione Maria Valença.

Monografia (Graduação) – Universidade Federal do Rio Grande do Norte.  
Centro de Ciências Exatas e da Terra. Departamento de Estatística.

1. Estatística – Monografia. 2. Estatísticas robustas – Monografia. 3. Modelo de regressão – Monografia. 4. Estatística gradiente – Monografia. 5. Comparação de testes de hipóteses – Monografia. I. Valença, Dione Maria. II. Título.

RN/UF/BSE-CCET

CDU: 519.2

Francisco Felipe de Queiroz

## **Estudo sobre má especificação na família de posição e escala**

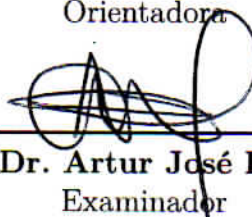
Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Aprovado em: 06 de Dezembro de 2016.



---

**Prof. Dr<sup>a</sup>. Dione Maria Valença**  
Orientadora



---

**Prof. Dr. Artur José Lemonte**  
Examinador



---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Mariana Correia de Araújo**  
Examinadora

Natal - RN

25 de Novembro de 2016

*À minha avó, Antônia, pelos ensinamentos.*

# Agradecimentos

A Deus, pelo dom da vida.

À minha mãe Lúcia, minha avó Antônia, e minha madrinha Nenzinha, pelo apoio e por sempre acreditarem em mim.

À professora Dione Maria Valença por ter aceitado ser minha orientadora neste trabalho, pelo apoio, paciência e pelos ensinamentos durante toda a minha graduação. Buscarei sempre seguir o seu exemplo de professora e educadora.

Aos professores do Departamento de Estatística, que estão sempre disponíveis e buscando ajudar nós alunos e, em especial: ao coordenador e professor 5 estrelas, Pledson Guedes, que está sempre buscando o melhor para os seus alunos e fazendo o possível para ajudá-los, aos professores André Pinho e Damião Nóbrega pelas conversas e conselhos durante o curso, aos professores Marcus Nunes, Carla Vivacqua, Iloneide Ramos e Paulo Roberto, pela disponibilidade em ajudar e apoio. Aos professores Carlos Dias, Liliane Gutierre e Viviane Klein do departamento de Matemática, pelo aprendizado e apoio nas disciplinas de Cálculo e Análise.

À professora Joanlise Marco, que me motivou a continuar no curso de Estatística.

Ao professor Artur Lemonte, pela disponibilidade em me ajudar e pelas orientações e conselhos.

À minha grande amiga Joelma, pelo amor à mim oferecido durante todos esses anos e por ter tornado esta etapa menos difícil. À minha amiga Maisa, pela paciência em me ouvir, pelos conselhos e pela amizade sincera. Aos amigos Fernando e Johnny, pelo companheirismo.

Aos meus amigos, que estiveram sempre presentes durante os últimos anos e, em especial, as minhas amigas Flávia e Vanessa, que me acompanham desde os primeiros meses de curso e que estão presentes em todos os momentos, ao meu amigo Pedro, pelo companheirismo e irmandade, a Karol e aos amigos PETianos: Inara, Joyce, Francimário, Erika, Lucas e Rodrigo. As tardes na UFRN não seriam as mesmas sem vocês!

Ao Programa de Educação Tutorial - PET, pelo suporte financeiro durante 3 anos da minha graduação e por todas as oportunidades que só acrescentaram na minha formação acadêmica.

*“Não sabia que era impossível. Foi lá e fez.”*  
*(Charlie Chaplin)*

# Resumo

A teoria da máxima verossimilhança é frequentemente utilizada na solução de problemas em diversas áreas da Estatística. Uma das suposições básicas da inferência baseada na teoria da verossimilhança é a de que o modelo escolhido para analisar os dados é, de fato, o modelo gerador dos dados e, quando isso não acontece, utilizar os procedimentos inferenciais usuais pode resultar em conclusões equivocadas. Nesse sentido, ainda que o problema de má especificação de um modelo paramétrico estatístico possa estar presente em diversos contextos, existem poucos trabalhos que tratam deste tema. Dessa forma, este trabalho dispõe-se a estudar os principais procedimentos inferenciais no contexto de má especificação de modelos e, em particular, em testes de hipóteses. São apresentadas versões robustas das estatísticas usuais para testar hipóteses quando o modelo é mal especificado, considerando modelos da família de posição e escala. Além disso, é estudado o desempenho destas estatísticas na presença de má especificação em modelos de regressão, considerando hipóteses simples e compostas. Finalmente, são realizados estudos de simulações de Monte Carlo. Nota-se que as versões robustas das estatísticas de testes apresentam bons resultados, tanto no contexto de amostras independentes e identicamente distribuídas, quanto no contexto de modelos de regressão.

**Palavras-chave:** Estatísticas robustas. Modelo de Regressão. Estatística gradiente. Comparação de testes de hipóteses.



# Abstract

The likelihood-based theory is often used in solving problems in various areas of statistics. One of the basic assumptions of inference based on this theory is that the chosen model correctly summarizes the data. When this does not happen, the usual procedures can result in misleading conclusions. In this sense, although the problem of misspecification can appear in several contexts, there are few papers dealing with this topic. Therefore, this work is designed to study the main inferential procedures in the context of misspecification and, in particular, hypotheses testing. Robust versions of the usual statistics are presented when the model is misspecified, considering location-scale family models. In addition, the performance of these statistics is studied in the presence of misspecification in regression models, considering simple and composite hypotheses and, finally, a simulation study is performed. It is noted that the robust versions of the statistics present good results, both in the context of independent and identically distributed samples as well as in the context of regression models.

**Keywords:** Robust statistics. Regression models. Gradient Statistic. Hypothesis test inference.

# Lista de tabelas

Tabela 1	– Taxa de rejeição para testar $\mathcal{H}_0: \theta = 1$ , com $n = 115$ . Considerando a distribuição correta $N(1, 1)$ e assumindo modelo <i>logístico</i> $(\theta, 1.5)$ . . .	43
Tabela 2	– Taxa de rejeição para testar $\mathcal{H}_0: \theta = 0$ com $n = 115$ . Considerando a distribuição correta <i>logística</i> $(0, 2)$ e assumindo modelo $t(2, \theta, 1)$ . . .	43
Tabela 3	– Taxa de rejeição para testar $\mathcal{H}_0: \theta = 1$ com $n = 115$ . Considerando a distribuição correta <i>logística</i> $(1, 3)$ e assumindo modelo $N(\theta, 1)$ . . . .	44
Tabela 4	– Taxa de rejeição para testar hipótese simples dada no Exemplo 5.2.1 com $n = 115$ para o modelo $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , considerando $Z_l \sim t(2, 0, 1)$ e assumindo $Z_l \sim \text{logística}(0, 1)$ , com $\sigma_0$ fixo e $x_{1l} = 1, \forall l$ .	47
Tabela 5	– Taxa de rejeição para testar hipótese $\mathcal{H}_0: \beta_2 = 0$ com $n = 115$ do modelo $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , considerando $Z_l \sim N(0, 1)$ e assumindo $Z_l \sim \text{logística}(0, 1)$ , com $\sigma_0$ fixo e $x_{1l} = 1, \forall l$ . . . . .	48
Tabela 6	– Taxa de rejeição para testar hipótese $\mathcal{H}_0: \beta_2 = 0$ com $n = 115$ do modelo $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \sigma_0 Z_l$ , considerando $Z_l \sim \text{logística}(0, 1)$ e assumindo $Z_l \sim t(2, 0, 1)$ , com $\sigma_0$ fixo e $x_{1l} = 1, \forall l$ . . . . .	48
Tabela 7	– Efeito do aumento do número de parâmetros nas estatísticas robustas para testar $\mathcal{H}_0: \beta_p = 0$ considerando o modelo $Y_l = \beta_1 x_{1l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l$ com $l = 1, \dots, n, Z_l \sim N(0, 1), n = 115, 200, 500$ , mas assumindo $Z_l \sim \text{logística}(0, 1)$ e $x_{1l} = 1, \forall l$ . . . . .	49
Tabela 8	– Efeito do aumento do número de parâmetros nas estatísticas usuais para testar $\mathcal{H}_0: \beta_p = 0$ considerando o modelo $Y_l = \beta_1 x_{1l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l$ com $l = 1, \dots, n, Z_l \sim N(0, 1), n = 115, 200, 500$ , mas assumindo $Z_l \sim \text{logística}(0, 1)$ e $x_{1l} = 1, \forall l$ . . . . .	54

# Sumário

	<b>Lista de tabelas</b> . . . . .	<b>10</b>
<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>13</b>
<b>1.1</b>	<b>Objetivos</b> . . . . .	<b>15</b>
<b>1.2</b>	<b>Organização dos Capítulos</b> . . . . .	<b>15</b>
<b>2</b>	<b>FAMÍLIA DE POSIÇÃO E ESCALA</b> . . . . .	<b>17</b>
<b>2.1</b>	<b>Exemplos de Membros da Família de Posição e Escala</b> . . . . .	<b>18</b>
<b>3</b>	<b>PROCEDIMENTOS INFERENCIAIS BASEADOS NA TEORIA DA VEROSSIMILHANÇA</b> . . . . .	<b>20</b>
<b>3.1</b>	<b>Modelo Corretamente Especificado</b> . . . . .	<b>20</b>
<b>3.1.1</b>	Estimação Pontual . . . . .	20
<b>3.1.2</b>	Teste de Hipóteses . . . . .	22
<b>3.2</b>	<b>Inferência sob Má especificação</b> . . . . .	<b>25</b>
<b>3.2.1</b>	Medidas de Informação e Interpretações . . . . .	26
<b>3.2.2</b>	Estimação Pontual . . . . .	28
<b>3.2.3</b>	Teste de Hipóteses sob Má Especificação . . . . .	29
<b>4</b>	<b>MÁ ESPECIFICAÇÃO NA FAMÍLIA DE POSIÇÃO E ESCALA</b> . . . . .	<b>33</b>
<b>4.1</b>	<b>Teste de Hipóteses</b> . . . . .	<b>33</b>
<b>4.1.1</b>	Hipótese Simples . . . . .	33
<b>4.1.2</b>	Hipótese Composta . . . . .	35
<b>5</b>	<b>MODELOS DE REGRESSÃO SOB MÁ ESPECIFICAÇÃO</b> . . . . .	<b>37</b>
<b>5.1</b>	<b>Modelos de Regressão e Má especificação</b> . . . . .	<b>37</b>
<b>5.2</b>	<b>Teste de Hipóteses</b> . . . . .	<b>40</b>
<b>6</b>	<b>SIMULAÇÕES</b> . . . . .	<b>42</b>
<b>6.1</b>	<b>Família de Posição e Escala</b> . . . . .	<b>42</b>
<b>6.2</b>	<b>Modelos de Regressão</b> . . . . .	<b>47</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>51</b>
	<b>APÊNDICE A – GRÁFICO DA DISCREPÂNCIA QUANTÍLICA RE- LATIVA</b> . . . . .	<b>53</b>

APÊNDICE B – EFEITO DO AUMENTO NO NÚMERO DE PARÂMETROS - ESTATÍSTICAS USUAIS . . . . .	54
REFERÊNCIAS . . . . .	55

# 1 Introdução

Inferência baseada na teoria da verossimilhança foi proposta inicialmente por Sir Ronald Fisher em 1912 e é amplamente utilizada em diversas áreas da estatística. Tal inferência é baseada em alguns princípios e uma das suposições feitas é que o modelo estatístico postulado aos dados esteja corretamente especificado (CORDEIRO, 1992). Na prática, consideramos que uma classe de modelos é apropriada para um conjunto de dados e um específico modelo é ajustado através de procedimentos inferenciais, ou seja, estimação e aplicação de testes apropriados sobre seus parâmetros. Feito isso, obtemos conclusões em nosso estudo com base neste ajuste. Contudo, os resultados inferenciais são fortemente influenciados pela correta escolha do modelo probabilístico para os dados, sendo que a má especificação do modelo pode levar a conclusões equivocadas. Nesse contexto, diz-se que o modelo é mal especificado.

Um dos principais procedimentos inferenciais que auxilia na tomada de decisão é o teste de hipótese. Os testes de hipóteses são utilizados quando o interesse é verificar se um ou mais parâmetros do modelo pertencem a uma dada região do espaço paramétrico. O primeiro exemplo de um teste de hipótese foi dado por Daniel Bernoulli em 1734 (CORDEIRO, 1992). As principais estatísticas de testes utilizadas para realização de testes de hipóteses, em casos regulares, são as estatísticas da razão de verossimilhanças (WILKS, 1938), de Wald (WALD, 1943), escore de Rao (RAO, 1948) e, recentemente, Terrell (2002) propôs a estatística gradiente, que vem sendo objeto de estudo em diferentes contextos. Estas estatísticas, no entanto, são adequadas quando o modelo postulado aos dados está corretamente especificado. Em particular, Lemonte (2013) apresenta um estudo de simulação utilizando estas estatísticas na presença de má especificação e os resultados mostram que a utilização delas pode causar sérios prejuízos na tomada de decisão.

White (1982) apresenta algumas das principais consequências quando técnicas usuais de verossimilhança são utilizadas na presença de má especificação. A distribuição assintótica do estimador de máxima verossimilhança é apresentada e é definido um teste baseado na matriz de informação que pode ser utilizado para a detecção de má especificação do modelo. Além disso, é apresentada a forma do estimador da variância assintótica do estimador de máxima verossimilhança na presença de má especificação, o chamado “estimador sanduíche”.

A distribuição assintótica da estatística da razão de verossimilhanças na presença de má especificação foi estudada em Kent (1982), admitindo que o modelo paramétrico postulado aos dados satisfaz algumas condições. Além disso, são analisadas algumas condições para as quais o teste da razão de verossimilhanças tem propriedades robustas,

ou seja, o teste continua tendo boas propriedades mesmo na presença de má especificação. Em Kent (1982) também são apresentadas as versões robustas das estatísticas de Wald e escore de Rao.

Com base nas propostas de Terrell (2002) e de Kent (1982), Lemonte (2013) apresenta uma versão robusta da estatística gradiente para testar hipóteses na presença de má especificação. Neste trabalho o autor considera diferentes modelos em um estudo de simulação para verificar o desempenho das estatísticas usuais e robustas quando o modelo é mal especificado e observa que a versão robusta da estatística gradiente aparenta ser bastante eficiente.

A versão robusta da estatística Wald, apresentada em Kent (1982) utiliza o “estimador sanduíche” em sua expressão. Este estimador é obtido pela expansão linear do vetor escore e reduz-se à inversa da matriz da informação, se a segunda identidade de Bartlett se verifica.

Um outro método para lidar com as consequências da má especificação é a modificação ou a correção da função de verossimilhança. Esta é a proposta abordada em Royall e Tsou (2003). Os autores estudam uma correção na função de verossimilhança para diminuir o efeito da má especificação do modelo, além de estudar o que ocorre com a informação esperada quando o modelo é mal especificado. Estes métodos não serão estudados neste trabalho, mas o estudo de Royall e Tsou (2003) é bastante útil para compreender o que ocorre, em termos de estimação, se ignorarmos o efeito da má especificação.

As principais referências neste tema, no entanto, consideram apenas o contexto de estimação para amostras independentes e identicamente distribuídas. Existem poucos trabalhos que abordam a situação em que a amostra não é identicamente distribuída, como ocorre em modelos de regressão, por exemplo. Siino, Fasola e Muggeo (2016) apresentam um estudo de simulação em modelos de regressão logística, em que é incluída a estatística gradiente robusta (LEMONTE, 2013), no contexto de pequenas amostras e dados espaçados, mas não estudam o desempenho desta estatística no contexto de má especificação. Nesse sentido, buscamos apresentar, neste trabalho, alguns estudos de simulação no contexto de modelos de regressão. Este, talvez, seja o principal resultado deste trabalho.

Ademais, estudamos uma importante família de distribuições, a família de posição e escala. Esta família de distribuições é amplamente utilizada em diversas áreas em virtude de suas boas propriedades e, em particular, em análise de sobrevivência. Portanto, concentramos nosso estudo em modelos pertencentes a esta família.

Finalmente, consideramos neste trabalho as estatísticas usuais, suas versões robustas e as estatísticas propostas recentemente (gradiente e gradiente robusta), a fim de verificar a eficiência destas no contexto de má especificação.

É importante mencionar que, neste trabalho, vamos usar, como notação de quanti-

dades numéricas, o ponto para separar unidades e decimais, concordando com a notação utilizada em publicações internacionais.

## 1.1 Objetivos

Os principais objetivos desta monografia são

- estudar as consequências da má especificação na inferência baseada na função de verossimilhança, com base nas ideias de White (1982) e Royall e Tsou (2003), além de explorar as principais referências nesta área;
- estudar o desempenho das estatísticas robustas propostas por Kent (1982) e Lemonte (2013) na família de posição e escala; e, finalmente,
- aplicar as ideias apresentadas em Kent (1982) e Lemonte (2013) em modelos de regressão.

Nesse sentido, os objetivos específicos podem ser resumidos em

- apresentar, através de uma revisão bibliográfica, alguns resultados importantes sobre as consequências da má especificação nos procedimentos usuais da inferência baseada na verossimilhança e apresentar as principais referências nesta área;
- reproduzir os resultados numéricos e os exemplos apresentados em Lemonte (2013);
- desenvolver as expressões das estatísticas robustas dadas em Lemonte (2013), considerando modelos da família de posição e escala;
- desenvolver expressões das estatísticas robustas no contexto de má especificação em modelos de regressão; e, por último,
- estudar, via simulações Monte Carlo, o desempenho das estatísticas robustas tanto no contexto da família de posição e escala (variáveis aleatórias independentes e identicamente distribuídas), quanto no contexto de modelos de regressão.

## 1.2 Organização dos Capítulos

Os capítulos deste trabalho são organizados da seguinte forma: No Capítulo 2 definimos a família de posição e escala e apresentamos alguns exemplos de modelos que fazem parte desta família, assim como algumas de suas propriedades. O Capítulo 3 é dividido em duas partes: na primeira parte, apresentamos uma breve revisão sobre métodos inferenciais baseados na função de verossimilhança, a fim de consolidar o conhecimento

sobre tais técnicas que serão de grande importância no decorrer deste trabalho e, no segundo momento, introduzimos o contexto de má especificação. Nesta segunda parte é feita uma revisão bibliográfica e um estudo das principais características e consequências da má especificação quando usamos o método da máxima verossimilhança para fazer inferência, tomando como base os artigos de White (1982) e Royall e Tsou (2003). Os Capítulos 4 e 5 apresentam os principais resultados deste trabalho. No Capítulo 4, apresentamos as expressões das estatísticas robustas para realização de testes de hipóteses no contexto de má especificação no caso de amostras aleatórias de um membro da família de posição e escala, considerando a hipótese nula simples e composta. Já no Capítulo 5, aplicamos os resultados analíticos desenvolvidos em Lemonte (2013) e Kent (1982) no contexto de modelos de regressão. Os resultados de simulação são apresentados no Capítulo 6. Consideramos, também, um breve estudo do poder dos testes (robustos e usuais) no contexto de má especificação. O Capítulo 7 é o último deste trabalho e nele apresentamos as principais conclusões e possíveis pesquisas futuras que podem ser realizadas neste tema.



## 2 Família de Posição e Escala

Seja  $X$  uma variável aleatória contínua. Dizemos que  $X$  pertence à família de posição e escala se sua função densidade de probabilidade é da forma

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} f_0\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R} \quad (2.1)$$

em que  $\mu \in \mathbb{R}$  é o parâmetro de posição,  $\sigma > 0$  é o parâmetro de escala e  $f_0(\cdot)$  é uma função densidade de probabilidade. A função  $f_0(\cdot)$  é frequentemente chamada de densidade padrão, já que ela é definida a partir da densidade  $f(\cdot)$  atribuindo valor 0 e 1 para os parâmetros de posição e escala, respectivamente.

A família de posição e escala possui propriedades bastante úteis. Algumas destas propriedades são resumidas no Teorema 2.0.1.

*Teorema 2.0.1.* Se  $X$  é uma variável aleatória contínua com função densidade de probabilidade  $f_X(x; \mu, \sigma)$  dada por (2.1) e  $Z$  é uma variável aleatória com função densidade de probabilidade  $f_0(z)$  e função distribuição acumulada  $F_0(z)$ , então

- a)  $F(x) = Pr(X \leq x) = F_0\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}$
- b)  $X = \mu + \sigma Z.$

Prova:

- a) *Pela definição da função de distribuição acumulada, segue que*

$$\begin{aligned} F(x) = Pr(X \leq x) &= \int_{-\infty}^x f_X(t; \mu, \sigma) dt = \int_{-\infty}^x \frac{1}{\sigma} f_0\left(\frac{t - \mu}{\sigma}\right) dt \\ &= \int_{-\infty}^{(x - \mu)/\sigma} f_0(u) du = F_0\left(\frac{x - \mu}{\sigma}\right). \end{aligned}$$

- b) *Considere uma variável aleatória  $W = \mu + \sigma Z$ . Temos*

$$F_W(w) = Pr(W \leq w) = Pr(\mu + \sigma Z \leq w) = Pr\left(Z \leq \frac{w - \mu}{\sigma}\right) = F_0\left(\frac{w - \mu}{\sigma}\right).$$

*Portanto,*

$$f_W(w) = \frac{\partial}{\partial w} F_W(w) = \frac{\partial}{\partial w} F_0\left(\frac{w - \mu}{\sigma}\right) = \frac{1}{\sigma} f_0\left(\frac{w - \mu}{\sigma}\right) = f_X(w).$$

*De modo que podemos escrever  $X = \mu + \sigma Z$ .*

Assim, diversos resultados associados a distribuição de  $X$  podem ser obtidos com base na distribuição de  $Z$ .

Considerando as transformações  $z_l = \sigma^{-1}(x_l - \mu)$  e  $m(z) = \log f_0(z)$ , a função de log-verossimilhança de  $(\mu, \sigma)$  dada uma amostra observada de tamanho  $n$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ , é expressa por

$$\ell(\mu, \sigma) = \sum_{l=1}^n m(z_l) - n \log \sigma.$$

Como podemos observar, a função de log-verossimilhança depende somente da densidade  $f_0(\cdot)$ , o que facilita na determinação das derivadas para os processos inferenciais, uma vez que  $f_0(\cdot)$  é, em geral, mais simples que a densidade  $f(\cdot)$ .

## 2.1 Exemplos de Membros da Família de Posição e Escala

Abaixo apresentaremos alguns exemplos de modelos que pertencem à família de posição e escala e que serão utilizados neste trabalho.

**Exemplo 2.1.1.** Distribuição valor extremo.

Considere o caso em que  $X \sim VE(\mu, \sigma)$ , com  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  e  $VE$  denotando uma distribuição valor extremo. Neste caso, temos que

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left[ \frac{x - \mu}{\sigma} - \exp \left( \frac{x - \mu}{\sigma} \right) \right], \quad x \in \mathbb{R}. \quad (2.2)$$

Note que  $f_X(x; \mu, \sigma)$  pode ser escrita como (2.1), com  $f_0(z) = \exp(z - e^z)$ , em que  $z \in \mathbb{R}$ . Assim,  $X = \mu + \sigma Z$ , com  $Z \sim VE(0,1)$ .

Além disso, considerando  $Y \sim Weibull(\alpha, \gamma)$ , podemos mostrar que  $\log Y \sim VE(\mu, \sigma)$ , em que  $\mu = \log \alpha$  e  $\sigma = 1/\gamma$ .

**Exemplo 2.1.2.** Distribuição logística.

Seja  $X \sim logística(\mu, \sigma)$ , em que  $\mu \in \mathbb{R}$  e  $\sigma > 0$ . Neste caso,

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left( \frac{x - \mu}{\sigma} \right) \left[ 1 + \exp \left( \frac{x - \mu}{\sigma} \right) \right]^{-2}, \quad x \in \mathbb{R}. \quad (2.3)$$

Como podemos observar, a densidade  $f_X(x; \mu, \sigma)$  pode ser representada na forma (2.1). Para isso, basta considerar  $f_0(z) = e^z (1 + e^z)^{-2}$ , com  $z \in \mathbb{R}$ , ou seja,  $X = \mu + \sigma Z$ , em que  $Z \sim logística(0,1)$ .

Considerando  $Y \sim log-logística(\alpha, \gamma)$ , podemos mostrar que  $\log Y \sim logística(\mu, \sigma)$ , em que  $\mu = \log \alpha$  e  $\sigma = 1/\gamma$ .

**Exemplo 2.1.3.** Distribuição normal.

Seja  $X \sim N(\mu, \sigma^2)$ , em que  $\mu \in \mathbb{R}$  e  $\sigma > 0$ . Neste caso,

$$f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}. \quad (2.4)$$

Podemos escrever a densidade  $f_X(x; \mu, \sigma)$  como apresentado em (2.1), considerando  $f_0(z) = (1/\sqrt{2\pi}) \exp(-z^2/2)$ , em que  $z \in \mathbb{R}$ . Desse modo, podemos escrever  $X = \mu + \sigma Z$ , com  $Z \sim N(0,1)$

Considerando  $Y \sim \text{lognormal}(\mu, \sigma)$ , podemos mostrar que  $\log Y \sim N(\mu, \sigma^2)$ .

**Exemplo 2.1.4.** Distribuição t-Student não padronizada.

Considere  $X \sim t\text{-Student}(\nu, \mu, \sigma)$ , em que  $\nu > 0$  representa os graus de liberdade,  $\mu \in \mathbb{R}$  é o parâmetro de posição e  $\sigma > 0$  o parâmetro de escala. Sendo assim,

$$f_X(x; \nu, \mu, \sigma) = \frac{\Gamma\left[\frac{\nu+1}{2}\right]}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu} \left(\frac{x - \mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}. \quad (2.5)$$

Não é difícil ver que  $f_X(x; \nu, \mu, \sigma)$  pode ser escrita como em (2.1), bastando, para isso, considerar

$$f_0(z; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Assim, podemos escrever  $X = \mu + \sigma Z$ , em que  $Z$  tem distribuição t-Student padrão.

A família de posição e escala tem uma grande aplicabilidade na área de análise de sobrevivência, na qual o principal interesse é modelar uma variável aleatória representando *o tempo até a ocorrência de um evento* (COLOSIMO; GIOLO, 2006). Nesse contexto, assumindo-se uma variável aleatória  $T$  para esse tempo, os principais modelos assumidos para  $T$  são tais que  $Y = \log T$  pertence à família de posição e escala. Desta forma, utilizar essa família para os procedimentos inferenciais é bem mais conveniente, dadas as propriedades apresentadas anteriormente. Lawless (2011) mostra que, nesses casos, conseguimos escrever a função de verossimilhança baseando-se, somente, na densidade padrão que, em geral, tem uma forma mais simples. Por exemplo, assumindo  $T \sim \text{Weibull}(\alpha, \gamma)$ , então  $Y = \log(T) \sim VE(\mu, \sigma)$ , com  $\mu = \log \alpha$  e  $\sigma = 1/\gamma$ .

Para mais detalhes e propriedades da família de posição e escala pode-se consultar Cordeiro (1992), Lawless (2011), entre outros.

## 3 Procedimentos Inferenciais Baseados na Teoria da Verossimilhança

Neste capítulo abordamos os principais resultados inferenciais (estimação pontual e teste de hipóteses) baseados na função de verossimilhança em dois aspectos: quando supomos que o modelo paramétrico escolhido para os dados é, de fato, o modelo gerador dos dados e quando estamos no contexto de má especificação, ou seja, quando o modelo escolhido como gerador dos dados é mal especificado.

### 3.1 Modelo Corretamente Especificado

Utilizando a notação de Lemonte (2013), considere  $\mathbf{X} = (X_1, \dots, X_n)^\top$  um vetor de variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade  $f(x; \boldsymbol{\theta})$ , que depende do vetor de parâmetros  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$  desconhecidos de dimensão  $p \times 1$ , com  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ , sendo  $\Theta$  o espaço paramétrico.

Associadas à densidade  $f$ , definimos a *função de verossimilhança* de  $\boldsymbol{\theta}$ , para a amostra observada de tamanho  $n$ ,  $\mathbf{x} = (x_1, \dots, x_n)^\top$ , como

$$L(\boldsymbol{\theta}) = \prod_{l=1}^n f(x_l; \boldsymbol{\theta}), \quad (3.1)$$

em que  $x_l$  é o valor observado da variável  $X_l$ . Note que a função de verossimilhança é a função de densidade conjunta interpretada como uma função de  $\boldsymbol{\theta}$  para  $\mathbf{x}$  fixado. A informação que iremos obter sobre  $\boldsymbol{\theta}$  será baseada na função de verossimilhança.

Em geral, trabalha-se com o logaritmo da função de verossimilhança, que é definido por

$$\ell_n(\boldsymbol{\theta}) = \sum_{l=1}^n \log f(x_l; \boldsymbol{\theta}). \quad (3.2)$$

#### 3.1.1 Estimação Pontual

A função de verossimilhança informa uma “ordem de preferência” entre os diversos valores de  $\boldsymbol{\theta}$ . Dada uma amostra observada, diremos que um valor de  $\boldsymbol{\theta}$ , digamos  $\boldsymbol{\theta}_1$ , é mais *verossível* que o outro, digamos  $\boldsymbol{\theta}_2$ , se a função de verossimilhança associada a  $\boldsymbol{\theta}_1$  for maior que a verossimilhança associada a  $\boldsymbol{\theta}_2$  (CORDEIRO, 1992). De maneira geral, entre os possíveis candidatos para estimar o verdadeiro valor de  $\boldsymbol{\theta}$  a partir do mesmo conjunto de dados, o vetor de parâmetros mais plausível é aquele de maior verossimilhança.

Sendo assim, o procedimento de estimação de máxima verossimilhança consiste em utilizar para estimar o vetor de parâmetros  $\boldsymbol{\theta}$  aquele vetor de parâmetros que maximiza a função de verossimilhança em  $\Theta$  (ou, analogamente, o vetor de parâmetros que maximiza a função de log-verossimilhança, dado que a função logaritmo é monótona e, portanto, maximizar  $L(\boldsymbol{\theta})$  e  $\ell_n(\boldsymbol{\theta})$  em  $\Theta$  são processos equivalentes). Para mais detalhes, ver Cordeiro (1992).

Portanto, define-se a estimativa de máxima verossimilhança de  $\boldsymbol{\theta}$  como o vetor  $\hat{\boldsymbol{\theta}} \in \Theta$  que maximiza  $\ell_n(\boldsymbol{\theta})$ , isto é,  $\ell_n(\hat{\boldsymbol{\theta}}) \geq \ell_n(\boldsymbol{\theta})$ ,  $\forall \boldsymbol{\theta} \in \Theta$ .

A primeira derivada da função de log-verossimilhança é chamada *função escore* (ou *vetor escore*) e é definido por

$$\mathbf{u}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i; \boldsymbol{\theta}), \quad (3.3)$$

de modo que  $\mathbf{u}_n(\boldsymbol{\theta})$  tenha a mesma dimensão de  $\boldsymbol{\theta}$ , isto é,  $p \times 1$ .

Nesse sentido, desde que a matriz de segundas derivadas do logaritmo da função de verossimilhança satisfaça algumas condições, o estimador de máxima verossimilhança é obtido, em casos regulares, pela solução do sistema de equações

$$\mathbf{u}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}.$$

Em geral, estas equações são não lineares e não fornecem soluções analíticas para  $\hat{\boldsymbol{\theta}}$ . No entanto, podemos encontrar  $\hat{\boldsymbol{\theta}}$  através do uso de procedimentos numéricos. Um dos principais métodos numéricos utilizados para a obtenção das estimativas de máxima verossimilhança é o método iterativo Newton-Raphson (CORDEIRO, 1999).

A *matriz da informação de Fisher* é definida por

$$\mathbf{K}_n(\boldsymbol{\theta}) = n\mathbf{K}(\boldsymbol{\theta}) = E[\mathbf{u}_n(\boldsymbol{\theta})\mathbf{u}_n(\boldsymbol{\theta})^\top] = \int_{-\infty}^{\infty} \mathbf{u}_n(\boldsymbol{\theta})\mathbf{u}_n(\boldsymbol{\theta})^\top f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x},$$

em que  $\mathbf{K}(\boldsymbol{\theta})$  denota a matriz da informação de Fisher com base em uma única observação.

Além disso, pode-se mostrar que

$$E[\mathbf{u}_n(\boldsymbol{\theta})] = \mathbf{0} \quad \text{e} \quad \text{Cov}[\mathbf{u}_n(\boldsymbol{\theta})] = \mathbf{K}_n(\boldsymbol{\theta}).$$

Sob algumas condições de regularidade, temos que

$$\mathbf{K}_n(\boldsymbol{\theta}) = -E[\partial \mathbf{u}_n(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^\top].$$

Estas condições de regularidade estão associadas às *identidades de Bartlett*, que são equações de grande importância na teoria da verossimilhança. Segundo Cordeiro (1999), as identidades mais conhecidas são:

$$k_r = 0 \quad \text{e} \quad k_{rs} + k_{r,s} = 0,$$

em que

$$k_r = E\left(\frac{\partial \ell_n}{\partial \theta_r}\right), \quad k_{rs} = E\left(\frac{\partial^2 \ell_n}{\partial \theta_r \partial \theta_s}\right), \quad k_{r,s} = E\left(\frac{\partial \ell_n}{\partial \theta_r} \frac{\partial \ell_n}{\partial \theta_s}\right),$$

$\ell_n$  denota a função log-verossimilhança e  $\theta_l$ ,  $l = 1, \dots, p$ , é o  $l$ -ésimo componente do vetor de parâmetros  $\boldsymbol{\theta}$ .

Os teoremas a seguir apresentam duas das principais propriedades do estimador de máxima verossimilhança quando o modelo está corretamente especificado. Estes teoremas são enunciados e provados em Casella e Berger (2011) e valem sob certas condições de regularidade.

*Teorema 3.1.1* (Normalidade Assintótica). Sob condições de regularidade e quando o tamanho da amostra é grande, temos que

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{K}(\boldsymbol{\theta})^{-1}),$$

em que  $\xrightarrow{d}$  representa convergência em distribuição.

O teorema enuncia que o estimador de máxima verossimilhança tem distribuição assintótica normal com média  $\boldsymbol{\theta}$  e matriz de variância e covariância assintótica dada por  $\mathbf{K}_n(\boldsymbol{\theta})^{-1}$ . Nesse sentido, temos que o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  é assintoticamente não viciado e sua variância assintótica coincide com o correspondente limite inferior das variâncias dos estimadores não viciados de  $\boldsymbol{\theta}$ . Portanto, em grandes amostras, temos que o estimador de máxima verossimilhança é eficiente (ver Bolfarine e Sandoval (2001) e Casella e Berger (2011)).

*Teorema 3.1.2.* Sejam  $\{\hat{\boldsymbol{\theta}}_n, n \geq 1\}$  uma sequência de estimadores. Dadas algumas condições,  $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$  quando  $n \rightarrow \infty$  e  $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$ , isto é,  $\hat{\boldsymbol{\theta}}_n$  converge em probabilidade para  $\boldsymbol{\theta}$ .

Em outras palavras, o Teorema diz que sequência de estimadores  $\{\hat{\boldsymbol{\theta}}_n, n \geq 1\}$  é consistente para  $\boldsymbol{\theta}$ .

### 3.1.2 Teste de Hipóteses

Existem muitas situações em que temos interesse em tomar a decisão de rejeitar ou não determinada afirmação baseando-se em um conjunto de evidências. Em Estatística, uma das principais formas de fazer isso é por meio dos *testes de hipóteses*. A teoria de testes de hipóteses é uma das partes principais da inferência baseada na verossimilhança e está diretamente relacionada com a teoria de estimação. Nesse contexto, o interesse principal consiste em determinar se um ou mais parâmetros pertencem a uma dada região do espaço paramétrico. Genericamente, uma *hipótese* é uma declaração sobre um parâmetro da população e os testes de hipóteses são compostos por duas hipóteses complementares: a *hipótese nula* e a *hipótese alternativa*, denotadas por  $\mathcal{H}_0$  e  $\mathcal{H}_1$ , respectivamente. Em um

problema de testes de hipóteses, depois de se observar a amostra, o experimentador deve decidir se rejeita ou não a hipótese nula.

Segundo Casella e Berger (2011), um teste de hipótese é uma regra que especifica:

- i. Para quais valores amostrais aceitamos  $\mathcal{H}_0$  como verdadeira.
- ii. Para quais valores amostrais  $\mathcal{H}_0$  é rejeitada e  $\mathcal{H}_1$  é aceita como verdadeira.

Utilizando a mesma notação de Cordeiro (1999), em testes paramétricos, as hipóteses são classificadas em *simples* e *compostas*. Dada uma distribuição que depende de  $p$  parâmetros e uma hipótese que especifica valores para  $d$  parâmetros, diremos que a hipótese é simples se  $d = p$  e composta se  $d < p$ . Podemos pensar, assim, que a hipótese simples “especifica completamente” a distribuição dos dados.

Seja  $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$  a função densidade conjunta dos dados, com  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ . Considere uma hipótese nula  $\mathcal{H}_0: \boldsymbol{\theta} \in \Theta_0 \subset \Theta$  versus a alternativa  $\mathcal{H}_1: \boldsymbol{\theta} \in \Theta_1 \subset \Theta$  ( $\Theta_1 = \Theta - \Theta_0$ ). O teste de hipóteses divide o espaço amostral (o conjunto de valores possíveis do vetor  $\mathbf{x}$ ) em duas regiões mutualmente excludentes:  $C$ , a região de rejeição de  $\mathcal{H}_0$ , chamada *região crítica*, e  $\bar{C}$ , a região complementar de não rejeição de  $\mathcal{H}_0$ . Em geral, nosso interesse é verificar se o vetor de dados pertence a  $C$  ou  $\bar{C}$ .

A região  $C$  é, em geral, determinada tal que, dado  $\mathcal{H}_0$ , a probabilidade de rejeitá-la (ou seja,  $\mathbf{X} \in C$ ) seja menor ou igual a um valor  $\alpha \in (0, 1)$  pré-especificado, isto é,

$$P(\mathbf{X} \in C \mid \boldsymbol{\theta} \in \Theta_0) \leq \alpha. \quad (3.4)$$

Em geral,  $\alpha$  é dito ser o nível de significância do teste e esta probabilidade é denominada probabilidade do *erro do tipo I*, que é a probabilidade de rejeitar a hipótese nula quando ela é verdadeira. Neste caso, a região  $C$  é determinada de modo que a probabilidade do erro de tipo I não exceda o valor pré-fixado  $\alpha$ . Outro erro que se pode cometer, o *erro do tipo II*, é o de não rejeitar a hipótese nula quando ela é falsa, sendo sua probabilidade dada por

$$\beta = P(\mathbf{X} \in \bar{C} \mid \boldsymbol{\theta} \in \Theta_1).$$

Como, em geral, pode-se encontrar várias regiões satisfazendo (3.4), deve-se decidir qual região crítica escolher. Este é o principal problema da teoria de testes de hipóteses. Nesse sentido, pode-se escolher uma região crítica  $C^*$  tal que ela maximize

$$1 - \beta = P(\mathbf{X} \in C \mid \boldsymbol{\theta} \in \Theta_1).$$

A probabilidade  $1 - \beta$ , para  $C$  fixo, como função do vetor de parâmetros  $\boldsymbol{\theta}$  especificado na hipótese alternativa, é denominada *função poder* do teste de  $\mathcal{H}_0$  versus  $\mathcal{H}_1$ .

Geralmente, um teste de hipóteses é especificado em termos de uma *estatística de teste*  $S(X_1, X_2, \dots, X_n) = S(\mathbf{X})$ , uma função da amostra. Por exemplo, um teste pode especificar que  $\mathcal{H}_0$  será rejeitada se  $\bar{X}$ , a média amostral, for maior que 10. Neste caso,  $S(\mathbf{X}) = \bar{X}$  é a estatística de teste e a região de rejeição é  $\{(x_1, x_2, \dots, x_n) : \bar{x} > 10\}$ .

Para mais detalhes a respeito da teoria de testes de hipóteses, tais como os critérios para escolha da região crítica, pode-se consultar Cordeiro (1999), Bolfarine e Sandoval (2001) e Casella e Berger (2011).

## Estatísticas Clássicas

### A) Hipóteses simples

Suponha que o interesse é testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$  contra a hipótese alternativa bilateral  $\mathcal{H}_1: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ . Neste caso, pode-se usar as estatísticas da razão de verossimilhanças (LR) (WILKS, 1938), de Wald (W) (WALD, 1943), escore de Rao (R) (RAO, 1948) e gradiente (T) (TERRELL, 2002) que são definidas, respectivamente, por

$$\begin{aligned} LR &= 2[\ell_n(\hat{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)], \\ W &= n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \\ R &= n^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \mathbf{K}(\boldsymbol{\theta}_0)^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0) \end{aligned}$$

e

$$T = \mathbf{u}_n(\boldsymbol{\theta}_0)^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

em que  $\hat{\boldsymbol{\theta}}$  é o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ .

Note que a estatística gradiente, proposta por Terrell (2002), não envolve em seu cálculo a matriz de informação (nem esperada, nem observada) sendo, portanto, mais simples de ser calculada do que as estatísticas  $W$  e  $R$ . Além disso, essa estatística envolve somente o produto do vetor escore avaliado sob  $\mathcal{H}_0$  e a diferença entre o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  restrito e irrestrito. Ainda, ela pode ser vista como uma combinação entre a estatística escore de Rao e a estatística de Wald. Em Lemonte (2016) é feito um estudo detalhado das propriedades da estatística gradiente em diferentes contextos.

Sob  $\mathcal{H}_0$  e para  $n$  grande, todas estas estatísticas possuem distribuição aproximadamente  $\chi_p^2$ . A hipótese nula é rejeitada se o valor observado da estatística do teste excede o quantil  $100(1 - \gamma)\%$  da distribuição  $\chi_p^2$ , em que  $\gamma = 1 - \alpha$  é o nível nominal do teste e  $\alpha$  é o nível de significância.

### B) Hipóteses Compostas

Considere a partição  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$ , em que  $\boldsymbol{\psi}$  e  $\boldsymbol{\lambda}$  tem dimensões  $q$  e  $p - q$ , respectivamente.



Suponha que o interesse é testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\psi} = \boldsymbol{\psi}_0$  contra a hipótese alternativa  $\mathcal{H}_1: \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$ , em que  $\boldsymbol{\psi}_0$  é um valor especificado e  $\boldsymbol{\lambda}$  é o vetor de parâmetros de perturbação. A partição de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  induz as partições

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{K}_{\psi\psi}(\boldsymbol{\theta}) & \mathbf{K}_{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{K}_{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{K}_{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \mathbf{K}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \mathbf{K}^{\psi\psi}(\boldsymbol{\theta}) & \mathbf{K}^{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{K}^{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{K}^{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix}.$$

Similarmente, temos a partição

$$\mathbf{u}_n(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{u}_{n\psi}(\boldsymbol{\theta}) \\ \mathbf{u}_{n\lambda}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^n \frac{\partial}{\partial \boldsymbol{\psi}} \log f(x_l; \boldsymbol{\theta}) \\ \sum_{l=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \log f(x_l; \boldsymbol{\theta}) \end{bmatrix}.$$

Sejam  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\lambda}}^\top)^\top$  e  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\psi}_0^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$ , respectivamente, os estimadores de máxima verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  irrestrito e o obtido sob a hipótese nula (restrito). Estes estimadores satisfazem  $\mathbf{u}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$  e  $\mathbf{u}_{n\lambda}(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ .

Neste caso, as estatísticas da razão de verossimilhanças, Wald, escore de Rao e gradiente, para testar a hipótese  $\mathcal{H}_0: \boldsymbol{\psi} = \boldsymbol{\psi}_0$  contra a hipótese alternativa  $\mathcal{H}_1: \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$  são dadas, respectivamente, por

$$LR = 2[\ell_n(\hat{\boldsymbol{\theta}}) - \ell_n(\tilde{\boldsymbol{\theta}})],$$

$$W = n(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)^\top \mathbf{K}^{\psi\psi}(\hat{\boldsymbol{\theta}})^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

$$R = n^{-1} \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top \mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}) \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})$$

e

$$T = \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0).$$

Sob a hipótese nula, a distribuição das estatísticas  $LR$ ,  $W$ ,  $R$  e  $T$  é aproximadamente qui-quadrado com  $q$  graus de liberdade, em que  $q$  representa a dimensão do vetor de parâmetros que estamos testando. Para mais detalhes, ver Cordeiro (1992) e Lemonte (2016).

## 3.2 Inferência sob Má especificação

Os testes apresentados na Seção 3.1.2 são adequadas quando o modelo escolhido para os dados é “bem especificado”. A presença de má especificação pode causar sérios prejuízos ao desempenho destas estatísticas. Nesta seção veremos as principais consequências e abordaremos as principais referências neste tema, entre elas: White (1982), Kent (1982), Lemonte (2013) e Royall e Tsou (2003). Neste trabalho, dedicamos nossa atenção às versões robustas das estatísticas de teste na presença de má especificação. A ideia principal

é propor, com base nas estatísticas de testes usuais, estatísticas de testes que tenham propriedades robustas.

Nesse contexto, assumimos que temos um vetor composto de  $n$  variáveis aleatórias independentes e identicamente distribuídas  $\mathbf{X} = (X_1, \dots, X_n)^\top$ , em que cada  $X_i$  tem função densidade de probabilidade desconhecida  $g(x)$ . Entretanto, iremos analisar a amostra observada com base no modelo paramétrico  $\{f(x; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ . Royall e Tsou (2003) estudam este problema e buscam responder perguntas como: O que acontece se a distribuição de  $X_i$  não é um membro deste modelo? Isto é, o que acontece se não existe um  $\boldsymbol{\theta}_0 \in \Theta$  para o qual  $g(x) = f(x; \boldsymbol{\theta}_0)$ ? Neste caso, o que a função de verossimilhança representa?

### 3.2.1 Medidas de Informação e Interpretações

Nesta seção introduzimos dois importantes critérios de informação que serão utilizados neste trabalho: o critério de informação de Kullback Leibler (KLIC) (KULLBACK; LEIBLER, 1951) e a informação de Fraser (FRASER et al., 1965).

Considere duas densidades  $f(x; \theta)$ , com  $\theta \in \Theta \subset \mathbb{R}$ , e  $g(x)$ . O critério de informação de Kullback-Leibler com respeito a um parâmetro  $\theta$  é definido por

$$I(\theta) = E_g \left\{ \log \left[ \frac{g(X)}{f(X; \theta)} \right] \right\} = \int \log[g(x)]g(x)dx - \int \log[f(x; \theta)]g(x)dx. \quad (3.5)$$

Note que os valores esperados são calculados com relação à verdadeira distribuição  $g(\cdot)$ . Intuitivamente,  $I(\theta)$  mensura a nossa “ignorância” sobre a verdadeira estrutura dos dados. Note que, quando a densidade  $g$  se aproxima da densidade  $f$ , o quociente  $g(x)/f(x; \theta)$  se aproxima de 1 e o  $\log [g(x)/f(x; \theta)]$  se aproxima de 0, de modo que  $I(\theta)$  se aproxima de 0, ou seja, nossa ignorância sobre o modelo gerador dos dados diminui.

A informação de Fraser de  $g : f$  é definida como

$$F(\theta) = E_g \{ \log[f(X; \theta)] \} = \int \log[f(x; \theta)]g(x)dx. \quad (3.6)$$

Note que minimizar KLIC implica maximizar a informação de Fraser dada por (3.6). Como a minimização de KLIC é com relação a  $\theta$ , e somente a segunda integral de (3.5) depende de  $\theta$ , então minimizar (3.5) equivale a maximizar (3.6).

Vale observar que

- $I(\theta) \geq 0, \forall \theta \in \Theta$ ;
- $I(\theta) = 0 \Leftrightarrow g(x) = f(x, \theta_0)$ , para algum  $\theta_0$ .

A primeira observação diz que  $E_g \{ \log[g(X)] \} \geq E_g \{ \log[f(X; \theta)] \}$  e, se pensarmos em termos de verossimilhança, já que podemos ver estas quantidades como sendo “versões

populacionais” das funções de log-verossimilhança de  $g$  e  $f$ , respectivamente, é intuitivo que a verossimilhança com respeito a verdadeira densidade seja maior do que aquela com densidade mal especificada.

A segunda observação decorre do fato de que se  $g(x) = f(x, \theta_0)$  para algum  $\theta_0$ , então os dois termos da integral em (3.5) apresentada anteriormente tornam-se iguais, de modo que  $I(\theta) = 0$ . Por outro lado, sendo  $I(\theta) = 0$ , podemos pensar, de maneira informal, que a nossa ignorância sobre a verdadeira estrutura dos dados é nula, ou seja, de alguma forma conhecemos o modelo gerador dos dados, de modo que  $g(x) = f(x, \theta_0)$  para algum  $\theta_0$ .

No contexto de má especificação, define-se  $\theta_g$  como o valor que minimiza (3.5) e maximiza (3.6), e este vetor de parâmetros é definido em Royall e Tsou (2003) como “objeto de inferência” uma vez que a verdadeira densidade é  $g(\cdot)$ .

Royall e Tsou (2003) verificam que, no contexto de má especificação, o objeto de inferência pode não ser igual ao “objeto de interesse”. Isto é, o vetor de parâmetros que estamos interessado em inferir pode não ser o vetor de parâmetros que estamos, de fato, inferindo. Para ilustrar, reproduzimos dois dos exemplos apresentados em Royall e Tsou (2003) em que a média é o objeto de interesse,  $E_g(X)$ .

### 1. Modelo Poisson ( $\theta$ ).

Temos que,

$$f(x; \theta) = \frac{\theta^x e^{-\theta}}{x!}, \theta > 0, x = 0, 1, \dots$$

Assim, a informação de Fraser  $E_g \{ \log[f(X; \theta)] \} = E_g(X) \ln(\theta) - \theta - E_g \{ \ln(X!) \}$  que (desde que exista  $E_g[\ln(X!)]$ ) é maximizada em  $\theta_g = E_g(X)$ . Aqui, portanto, o objeto de inferência é o objeto de interesse.

### 2. Modelo lognormal ( $\mu, \sigma^2$ ), com $\sigma$ fixo.

Reparametrizando a densidade em termos do valor esperado  $\theta = \exp(\mu + \sigma^2/2)$ , temos que

$$f(x; \theta) = (2\pi\sigma^2 x^2)^{-1/2} \exp \left\{ -\frac{[\ln(x) - \ln(\theta) + \sigma^2/2]^2}{2\sigma^2} \right\}, \theta > 0, x > 0.$$

A informação de Fraser  $E_g \{ \log[f(X; \theta)] \}$  é maximizada quando  $\ln(\theta) - \sigma^2/2 = E_g[\ln(X)]$ , então  $\theta_g = \exp\{E_g[\ln(X)] + \sigma^2/2\}$ .

Para o primeiro modelo, o objeto de inferência,  $\theta_g$ , é igual ao objeto de interesse,  $E_g(X)$ , quando o modelo falha, ou seja, quando o modelo não é Poisson. No segundo caso, quando a verdadeira distribuição  $g$  não é lognormal, então a função de verossimilhança não representa evidência sobre a média,  $E_g(X)$ , mas sobre a quantidade  $\theta_g = \exp\{E_g[\ln(X)] + \sigma^2/2\}$ .

### 3.2.2 Estimação Pontual

A função de verossimilhança de  $\boldsymbol{\theta}$  e o logaritmo da função de verossimilhança para uma amostra observada  $\mathbf{x} = (x_1, \dots, x_n)^\top$  são dados por (3.1) e (3.2), respectivamente. Aqui, chamamos a atenção para o fato de que a função de verossimilhança de  $\boldsymbol{\theta}$  é definida usando a densidade que assumimos e não a verdadeira densidade  $g(\cdot)$ . Analogamente, o estimador de máxima verossimilhança (EMV) é o vetor de parâmetros  $\hat{\boldsymbol{\theta}} \in \Theta$  tal que

$$\ell_n(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} \ell_n(\boldsymbol{\theta}).$$

Sob má especificação de modelos, White (1982) define o logaritmo da função de verossimilhança como *função de log quase verossimilhança*. Isto é feito pois, nas funções de verossimilhança usuais, especificamos uma família de distribuições para os dados, o que não ocorre no contexto de má especificação. De fato, a função quase-verossimilhança não exige que especifiquemos a família de distribuições corretas para os dados, mas apenas que conheçamos as relações funcionais entre os seus dois primeiros momentos (CORDEIRO, 1992). Analogamente, o estimador de máxima verossimilhança é chamado de *estimador de máxima quase-verossimilhança*. No entanto, a fim de simplificar as notações aqui utilizadas, omitiremos este fato. Para mais detalhes com relação as suposições utilizadas para a fundamentação das definições aqui apresentadas, sugerimos a leitura de White (1982).

Sabemos que quando  $f$  contém a verdadeira estrutura  $g$ , isto é,

$$g(x) = f(x; \boldsymbol{\theta}_0), \quad (3.7)$$

o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  é consistente para  $\boldsymbol{\theta}_0$ , sob algumas condições de regularidade (ver Teorema 3.1.2).

Se (3.7) não ocorre, Akaike (1973) nota que  $\hat{\boldsymbol{\theta}}$  é um estimador natural para  $\boldsymbol{\theta}_g$ , o vetor de parâmetros que minimiza o critério de informação de Kullback-Leibler, definido na Seção 3.2.1. Sendo assim, a função de verossimilhança representa evidências sobre  $\boldsymbol{\theta}_g$ .

Os resultados a seguir são apresentados em White (1982) e enunciam algumas das propriedades do estimador de máxima verossimilhança no contexto de má especificação de modelos.

*Teorema 3.2.1 (Consistência).* Sejam  $\{\hat{\boldsymbol{\theta}}_n, n \geq 1\}$  uma sequência de estimadores. Dadas algumas condições,  $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_g$  quando  $n \rightarrow \infty$ , isto é,  $\hat{\boldsymbol{\theta}}_n$  converge em probabilidade para  $\boldsymbol{\theta}_g$ .

Com isso, notamos que a sequência de estimadores é consistente para  $\boldsymbol{\theta}_g$ , o vetor de parâmetros que minimiza KLIC. Nesse contexto, o estimador de máxima verossimilhança converge para o vetor de parâmetros que minimiza a nossa ignorância sobre a verdadeira estrutura dos dados. Ou seja, estamos “minimizando nossa ignorância” sobre a verdadeira

estrutura dos dados. O teorema a seguir apresenta a distribuição assintótica do estimador de máxima verossimilhança no contexto de má especificação de modelos.

*Teorema 3.2.2* (Normalidade Assintótica). Dadas algumas suposições,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_g) \xrightarrow{d} N(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}_g)),$$

em que a matriz  $\mathbf{C}$  é definida como

$$\mathbf{C}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})^{-1} \mathbf{J}(\boldsymbol{\theta}) \mathbf{H}(\boldsymbol{\theta})^{-1},$$

em que

$$\mathbf{J}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \mathbf{u}(\boldsymbol{\theta}) \mathbf{u}(\boldsymbol{\theta})^{\top} g(x) dx, \quad \mathbf{H}(\boldsymbol{\theta}) = - \int_{-\infty}^{\infty} \frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}} g(x) dx,$$

com

$$\mathbf{u}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x; \boldsymbol{\theta}) = \left( \frac{\partial}{\partial \theta_j} \log f(x; \boldsymbol{\theta}) \right)_{j=1, \dots, p}.$$

Note que a matriz  $\mathbf{J}(\boldsymbol{\theta})$  é a covariância da função escore e a matriz  $\mathbf{H}(\boldsymbol{\theta})$  é o valor esperado da derivada da função escore. Ainda, perceba que  $\mathbf{J}(\boldsymbol{\theta}) \neq \mathbf{H}(\boldsymbol{\theta})$ , ou seja, a segunda identidade de Bartlett não se verifica. Se o modelo assumido é correto, isto é,  $g(x) = f(x; \boldsymbol{\theta}_0)$  para algum  $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ , então o Teorema 3.1.1, que fala da normalidade assintótica do estimador de máxima verossimilhança, é um caso particular do Teorema 3.2.2. Isso por que a expressão  $\mathbf{H}(\boldsymbol{\theta}_g)^{-1} \mathbf{J}(\boldsymbol{\theta}_g) \mathbf{H}(\boldsymbol{\theta}_g)^{-1}$  reduz-se a inversa da matriz da informação de Fisher, se as identidades de Bartlett se verificam.

Royall e Tsou (2003) propõem um ajuste na função de verossimilhança tornando-a robusta na presença de má especificação de modelos. Uma das condições para que isso ocorra é que o objeto de interesse seja igual ao objeto de inferência, este aspecto não será estudado neste trabalho. Aqui, consideramos que, embora não conheçamos a verdadeira estrutura dos dados, isto é,  $g$ , para que os testes robustos apresentem bons resultados, os valores esperados dos modelos  $g$  e  $f$  devem ser iguais.

### 3.2.3 Teste de Hipóteses sob Má Especificação

#### A) Hipóteses Simples

Suponha, inicialmente, que o interesse é testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\theta}_g = \boldsymbol{\theta}_0$  contra a hipótese alternativa  $\mathcal{H}_1: \boldsymbol{\theta}_g \neq \boldsymbol{\theta}_0$ .

No caso de modelo mal especificado, a versão robusta da estatística de Wald (KENT, 1982) é dada por

$$S_W = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\top} \mathbf{H}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad (3.8)$$

e a versão robusta da estatística escore de Rao (KENT, 1982) é dada por

$$S_R = n^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)^{\top} \mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0), \quad (3.9)$$

em que  $\mathbf{u}_n(\boldsymbol{\theta})$  é dado em (3.3). Segundo Kent (1982), as estatísticas robustas  $S_W$  e  $S_R$  tem distribuição assintótica  $\chi_p^2$  sob a hipótese nula, e não existe uma versão robusta para a estatística da razão de verossimilhanças que tenha distribuição assintótica qui-quadrado.

Recentemente, Lemonte (2013) propôs uma versão robusta da estatística gradiente para testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\theta}_g = \boldsymbol{\theta}_0$  versus  $\mathcal{H}_1: \boldsymbol{\theta}_g \neq \boldsymbol{\theta}_0$ . A estatística gradiente robusta é dada por

$$S_T = \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad (3.10)$$

e, sob a hipótese nula, o autor mostra que  $S_T$  também tem distribuição assintótica  $\chi_p^2$ .

Como, em geral,  $\mathbf{H}(\boldsymbol{\theta})$  e  $\mathbf{J}(\boldsymbol{\theta})$  são desconhecidos, pode-se usar as quantidades  $\hat{\mathbf{H}}_n(\boldsymbol{\theta})$  e  $\hat{\mathbf{J}}_n(\boldsymbol{\theta})$  que, quando avaliadas sob o estimador de máxima verossimilhança, são estimadores consistentes (KENT, 1982), e são definidos como

$$\hat{\mathbf{H}}_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{l=1}^n \frac{\partial \mathbf{u}^{(l)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}, \quad (3.11)$$

$$\hat{\mathbf{J}}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{l=1}^n \mathbf{u}^{(l)}(\boldsymbol{\theta}) \mathbf{u}^{(l)}(\boldsymbol{\theta})^\top, \quad (3.12)$$

em que

$$\mathbf{u}^{(l)}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_l; \boldsymbol{\theta}) = \left( \frac{\partial}{\partial \theta_j} \log f(x_l; \boldsymbol{\theta}) \right)_{j=1, \dots, p}, \quad l = 1, \dots, n$$

e

$$\mathbf{u}_n(\boldsymbol{\theta}) = \sum_{l=1}^n \mathbf{u}^{(l)}(\boldsymbol{\theta}).$$

Assim, as estatísticas robustas  $S_W^*$ ,  $S_R^*$  e  $S_T^*$  são obtidas substituindo  $\mathbf{H}$  e  $\mathbf{J}$  por seus estimadores em (3.8), (3.9) e (3.10), e são da forma

$$S_W^* = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \hat{\mathbf{H}}_n(\boldsymbol{\theta}_0) \hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)^{-1} \hat{\mathbf{H}}_n(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

$$S_R^* = n^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)$$

e

$$S_T^* = \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)^{-1} \hat{\mathbf{H}}_n(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Kent (1982) e Lemonte (2013) mostram que, sob a hipótese nula,  $S_W^*$ ,  $S_R^*$  e  $S_T^*$  continuam tendo distribuição aproximadamente qui-quadrado com  $p$  graus de liberdade, para  $n$  grande.

A quantidade  $\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}}) \hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}}) \hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})$  é conhecida como *estimador sanduíche* e é um estimador consistente para  $\mathbf{C}(\boldsymbol{\theta}_g)$ , a variância assintótica de  $\hat{\boldsymbol{\theta}}$ , apresentada no Teorema 3.2.2 (LEMONTE, 2013).

### B) Hipóteses Compostas

Considere a partição  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  e, da mesma forma,  $\boldsymbol{\theta}_g = (\boldsymbol{\psi}_g^\top, \boldsymbol{\lambda}_g^\top)^\top$ , em que  $\boldsymbol{\psi}$  e  $\boldsymbol{\lambda}$  tem dimensão  $q$  e  $p - q$ , respectivamente. Suponha que o interesse seja testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\psi}_g = \boldsymbol{\psi}_0$  contra a hipótese alternativa  $\mathcal{H}_1: \boldsymbol{\psi}_g \neq \boldsymbol{\psi}_0$ , em que  $\boldsymbol{\psi}_0$  é um valor especificado de  $\boldsymbol{\psi}$ , e  $\boldsymbol{\lambda}$  é o vetor de parâmetros de perturbação. A partição de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  induz as partições das matrizes:

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}_{\psi\psi}(\boldsymbol{\theta}) & \mathbf{H}_{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{H}_{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{H}_{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix}, \quad (3.13)$$

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{\psi\psi}(\boldsymbol{\theta}) & \mathbf{J}_{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{J}_{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{J}_{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix}, \quad (3.14)$$

$$\mathbf{H}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \mathbf{H}^{\psi\psi}(\boldsymbol{\theta}) & \mathbf{H}^{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{H}^{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{H}^{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \mathbf{J}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \mathbf{J}^{\psi\psi}(\boldsymbol{\theta}) & \mathbf{J}^{\psi\lambda}(\boldsymbol{\theta}) \\ \mathbf{J}^{\lambda\psi}(\boldsymbol{\theta}) & \mathbf{J}^{\lambda\lambda}(\boldsymbol{\theta}) \end{bmatrix}.$$

Similarmente, temos a partição

$$\mathbf{u}_n(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{u}_{n\psi}(\boldsymbol{\theta}) \\ \mathbf{u}_{n\lambda}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^n \frac{\partial}{\partial \boldsymbol{\psi}} \log f(x_l; \boldsymbol{\theta}) \\ \sum_{l=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \log f(x_l; \boldsymbol{\theta}) \end{bmatrix}. \quad (3.15)$$

Sejam  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\lambda}}^\top)^\top$  e  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\psi}_0^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$ , respectivamente, os estimadores de máxima verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  irrestrito e obtido sob a hipótese nula (restrito). Estes estimadores satisfazem  $\mathbf{u}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$  e  $\mathbf{u}_{n\lambda}(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ .

Segundo Kent (1982), as estatísticas Wald robusta e escore robusta são dadas, respectivamente, por

$$S_W = n(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)^\top [\mathbf{H}(\tilde{\boldsymbol{\theta}}) \mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\psi\psi} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$$

e

$$S_R = n^{-1} \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top \mathbf{H}^{\psi\psi}(\tilde{\boldsymbol{\theta}}) [\mathbf{H}(\tilde{\boldsymbol{\theta}}) \mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\psi\psi} \mathbf{H}^{\psi\psi}(\tilde{\boldsymbol{\theta}}) \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}}),$$

em que  $[\mathbf{A}]_{\psi\psi}$  denota uma matriz de dimensão  $q \times q$  relacionada ao vetor de parâmetros de interesse  $\boldsymbol{\psi}$ , obtida de  $\mathbf{A}$ . A versão robusta da estatística gradiente (LEMONTE, 2013) passa a ser dada pela expressão

$$S_T = \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top [\mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\psi\psi} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0).$$

Similarmente, podemos escrever  $S_W$ ,  $S_R$  e  $S_T$  da forma

$$S_W = n(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)^\top [\mathbf{H}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{J}(\tilde{\boldsymbol{\theta}}) \mathbf{H}(\tilde{\boldsymbol{\theta}})^{-1}]_{\psi\psi}^{-1} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

$$S_R = n^{-1} \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top \mathbf{H}^{\psi\psi}(\tilde{\boldsymbol{\theta}}) [\mathbf{H}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{J}(\tilde{\boldsymbol{\theta}}) \mathbf{H}(\tilde{\boldsymbol{\theta}})^{-1}]_{\psi\psi}^{-1} \mathbf{H}^{\psi\psi}(\tilde{\boldsymbol{\theta}}) \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}}),$$

$$S_T = \mathbf{u}_{n\psi}(\tilde{\boldsymbol{\theta}})^\top [\mathbf{H}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{J}(\tilde{\boldsymbol{\theta}})]_{\psi\psi}^{-1} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0).$$

De acordo com Kent (1982) e Lemonte (2013), sob a hipótese nula, o limite em distribuição das estatísticas  $S_W$ ,  $S_R$  e  $S_T$  é qui-quadrado com  $q$  graus de liberdade, em que  $q$  representa a dimensão do vetor de parâmetros que estamos testando. Sugerimos a leitura de Kent (1982) para detalhes da distribuição das estatísticas  $S_W$  e  $S_R$ , e a leitura de Lemonte (2013) e Lemonte (2016) no que se refere a estatística  $S_T$ .

Como, na prática, as quantidades  $\mathbf{H}(\boldsymbol{\theta})$  e  $\mathbf{J}(\boldsymbol{\theta})$  são desconhecidas, pode-se utilizar os seus respectivos estimadores empíricos  $\hat{\mathbf{H}}_n(\boldsymbol{\theta})$  e  $\hat{\mathbf{J}}_n(\boldsymbol{\theta})$ . A hipótese nula é rejeitada se o valor observado da estatística de teste excede o quantil  $100(1 - \gamma)\%$  da distribuição  $\chi_q^2$ .



## 4 Má Especificação na Família de Posição e Escala

Neste capítulo, estudamos o comportamento das estatísticas robustas em modelos pertencentes à família de posição e escala. Neste sentido, abordamos inicialmente o caso em que o nosso interesse é testar hipóteses simples e, para tanto, consideramos o parâmetro de escala fixo. Alguns exemplos são apresentados para as principais distribuições que pertencem a esta família. Em seguida, é apresentado o caso em que desejamos testar hipóteses sobre uma partição do vetor de parâmetros (hipóteses compostas). As formas das estatísticas robustas são apresentadas e são citadas algumas referências que abordam este caso com mais detalhes.

### 4.1 Teste de Hipóteses

#### 4.1.1 Hipótese Simples

Assumimos que temos um vetor composto de  $n$  variáveis aleatórias independentes e identicamente distribuídas  $\mathbf{X} = (X_1, \dots, X_n)^\top$  em que cada  $X_l$  tem função densidade de probabilidade desconhecida  $g(x)$ . No entanto, suponha que iremos analisar estes dados utilizando o modelo paramétrico  $f(x; \theta)$ , com  $\theta \in \Theta$ . Considere, ainda, que a densidade  $f(x; \theta)$  pode ser escrita como em (2.1), em que  $\theta \in \mathbb{R}$  é o parâmetro de posição,  $f_0(\cdot)$  é uma função densidade de probabilidade e  $\sigma_0 > 0$  é o parâmetro de escala, que consideramos fixo. Ou seja, estamos assumindo que o modelo escolhido para sumarizar os dados pertence à família de posição e escala.

Nesse contexto, considerando  $z_l = \sigma_0^{-1}(x_l - \theta)$ , a função de log-verossimilhança de  $\theta$  com relação a amostra observada  $\mathbf{x} = (x_1, \dots, x_n)^\top$  é dada por

$$\ell_n(\theta) = \ell_n(\theta; \mathbf{x}) = \sum_{l=1}^n \log f(x_l; \theta) = -n \log \sigma_0 + \sum_{l=1}^n \log f_0(z_l).$$

Desta forma, usando a notação  $m(z) = \log f_0(z)$ , pode-se mostrar que

$$u^{(l)}(\theta) = \frac{\partial}{\partial \theta} \log f(x_l; \theta) = -\frac{1}{\sigma_0} m'(z_l),$$

e, portanto,

$$u_n(\theta) = \sum_{l=1}^n u^{(l)}(\theta) = -\frac{1}{\sigma_0} \sum_{l=1}^n m'(z_l),$$

em que  $m'(z_l) = \frac{\partial m(z_l)}{\partial z_l}$  e  $m''(z_l) = \frac{\partial^2 m(z_l)}{\partial z_l^2}$ . Utilizando os estimadores empíricos de  $H(\theta)$  e  $J(\theta)$  dados em (3.11) e (3.12), respectivamente, temos que

$$\hat{H}_n(\theta) = -\frac{1}{n\sigma_0^2} \sum_{l=1}^n m''(z_l)$$

e

$$\hat{J}_n(\theta) = \frac{1}{n\sigma_0^2} \sum_{l=1}^n [m'(z_l)]^2.$$

Suponha que desejamos testar a hipótese  $\mathcal{H}_0: \theta_g = \theta_0$  contra a hipótese alternativa  $\mathcal{H}_1: \theta_g \neq \theta_0$ . Segue que

$$S_W^* = \left( \frac{\hat{\theta} - \theta_0}{\sigma_0} \right)^2 \frac{\left[ \sum_{l=1}^n m''(z_l) \right]^2}{\sum_{l=1}^n [m'(z_l)]^2}, \quad (4.1)$$

$$S_R^* = \frac{\left[ \sum_{l=1}^n m'(z_l) \right]^2}{\sum_{l=1}^n [m'(z_l)]^2} \quad (4.2)$$

e

$$S_T^* = \left( \frac{\hat{\theta} - \theta_0}{\sigma_0} \right) \frac{\left[ \sum_{l=1}^n m'(z_l) \right] \left[ \sum_{l=1}^n m''(z_l) \right]}{\sum_{l=1}^n [m'(z_l)]^2}. \quad (4.3)$$

Assim, basta conhecermos a função  $m(\cdot)$  para determinarmos a forma destas estatísticas para modelos pertencentes a família de posição e escala.

Apresentamos, a seguir, as expressões das estatísticas robustas para alguns modelos da família de posição e escala. Em todos os casos consideramos  $X_l$  com distribuição pertencente à família (2.1), com parâmetro de posição  $\theta$  e parâmetro de escala  $\sigma_0$  (conhecido). Nosso interesse é testar  $\mathcal{H}_0: \theta = \theta_0$  contra a hipótese alternativa  $\mathcal{H}_1: \theta \neq \theta_0$ .

**Exemplo 4.1.1.** Distribuição valor extremo.

Se  $X_l, l = 1, \dots, n$ , são variáveis aleatórias independentes e identicamente distribuídas seguindo distribuição valor extremo com parâmetros  $\theta$  e  $\sigma_0$ , então as expressões das estatísticas robustas para testar as hipóteses de interesse são dadas por

$$S_W^* = \left( \frac{\hat{\theta} - \theta_0}{\sigma_0} \right)^2 \frac{\left[ \sum_{l=1}^n e^{z_l} \right]^2}{\sum_{l=1}^n [1 - e^{z_l}]^2},$$

$$S_R^* = \frac{\left[ \sum_{l=1}^n (1 - e^{z_l}) \right]^2}{\sum_{l=1}^n [1 - e^{z_l}]^2}$$

e

$$S_T^* = \left( \frac{\theta_0 - \hat{\theta}}{\sigma_0} \right) \frac{\left[ \sum_{l=1}^n (1 - e^{z_l}) \right] \left[ \sum_{l=1}^n e^{z_l} \right]}{\sum_{l=1}^n [1 - e^{z_l}]^2}.$$

**Exemplo 4.1.2.** Distribuição logística.

Considere, agora, que temos um conjunto de observações provenientes de uma variável aleatória com distribuição logística com parâmetros  $\theta$  e  $\sigma_0$ , cuja densidade é apresentada em (2.3). Nesse contexto, as expressões das estatísticas robustas são dadas por (4.1), (4.2) e (4.3), em que

$$m'(z_l) = \frac{1 - e^{z_l}}{1 + e^{z_l}} \quad \text{e} \quad m''(z_l) = -\frac{2e^{z_l}}{(1 + e^{z_l})^2}.$$

**Exemplo 4.1.3.** Distribuição normal.

Suponha que  $X_l, l = 1, \dots, n$ , são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal de parâmetros  $\theta$  e  $\sigma_0$ , cuja distribuição é apresentada em (2.4). Neste caso, as expressões das estatísticas robustas são dadas por

$$S_W^* = S_R^* = S_T^* = \frac{\left[ \sum_{l=1}^n z_l \right]^2}{\sum_{l=1}^n z_l^2}.$$

**Exemplo 4.1.4.** Distribuição t-Student não padronizada.

Sejam  $X_l, l = 1, \dots, n$ , variáveis aleatórias independentes e identicamente distribuídas com distribuição t-Student de parâmetros  $\theta$ ,  $\sigma_0$  e  $\nu$ , cuja distribuição é apresentada em (2.5). Neste caso, as expressões das estatísticas robustas são dadas por (4.1), (4.2) e (4.3), em que

$$m'(z_l) = -\frac{(v+1)z_l}{v+z_l^2} \quad \text{e} \quad m''(z_l) = \frac{(v+1)(z_l^2 - v)}{(v+z_l^2)^2}.$$

## 4.1.2 Hipótese Composta

Assuma que temos um vetor composto de  $n$  variáveis aleatórias independentes e identicamente distribuídas  $\mathbf{X} = (X_1, \dots, X_n)^\top$  em que cada  $X_l$  tem função densidade de probabilidade desconhecida  $g(x)$ . Suponha que iremos analisar estes dados utilizando o modelo paramétrico  $f(x; \boldsymbol{\theta})$ , com  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ . Considere, ainda, que a densidade  $f(x; \boldsymbol{\theta})$  pode ser escrito como (2.1), com  $\boldsymbol{\theta} = (\mu, \sigma)^\top$ , em que  $\mu \in \mathbb{R}$  é o parâmetro de posição e  $\sigma > 0$  é o parâmetro de escala.

Com isso, defini-se o logarítmo da função de verossimilhança para  $\boldsymbol{\theta} = (\mu, \sigma)^\top$  da forma

$$\ell_n(\boldsymbol{\theta}) = \sum_{l=1}^n \log f_0(z_l) - n \log \sigma.$$

Neste caso, temos que

$$\mathbf{u}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{l=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_l; \boldsymbol{\theta}),$$

e, fazendo  $m(z) = \log f_0(z)$ , segue que

$$u_{n\mu}(\boldsymbol{\theta}) = -\frac{1}{\sigma} \sum_{l=1}^n m'(z_l)$$

e

$$u_{n\sigma}(\boldsymbol{\theta}) = -\frac{1}{\sigma} \sum_{l=1}^n [1 + z_l m'(z_l)].$$

Considere que o interesse seja testar as hipóteses  $\mathcal{H}_0: \mu = \mu_0$  contra  $\mathcal{H}_1: \mu \neq \mu_0$ , sendo  $\sigma$  o parâmetro de perturbação. Neste caso, as matrizes definidas em (3.11) e (3.12), são dadas por

$$\hat{\mathbf{H}}_n(\boldsymbol{\theta}) = \begin{bmatrix} \hat{H}_{\mu\mu}(\boldsymbol{\theta}) & \hat{H}_{\mu\sigma}(\boldsymbol{\theta}) \\ \hat{H}_{\sigma\mu}(\boldsymbol{\theta}) & \hat{H}_{\sigma\sigma}(\boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \hat{\mathbf{J}}_n(\boldsymbol{\theta}) = \begin{bmatrix} \hat{J}_{\mu\mu}(\boldsymbol{\theta}) & \hat{J}_{\mu\sigma}(\boldsymbol{\theta}) \\ \hat{J}_{\sigma\mu}(\boldsymbol{\theta}) & \hat{J}_{\sigma\sigma}(\boldsymbol{\theta}) \end{bmatrix},$$

cujos elementos são

$$\begin{aligned} \hat{H}_{\mu\mu}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} \sum_{l=1}^n m''(z_l), & \hat{H}_{\sigma\sigma}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} \sum_{l=1}^n [1 + 2z_l m'(z_l) + z_l^2 m''(z_l)], \\ \hat{H}_{\sigma\mu}(\boldsymbol{\theta}) &= \hat{H}_{\mu\sigma}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{l=1}^n [m'(z_l) + z_l m''(z_l)], & \hat{J}_{\mu\mu}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} [m'(z_l)]^2, \\ \hat{J}_{\sigma\sigma}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2} [1 + z_l m'(z_l)]^2 & \text{e} \quad \hat{J}_{\sigma\mu}(\boldsymbol{\theta}) &= \hat{J}_{\mu\sigma}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \{m'(z_l) + z_l [m'(z_l)]^2\}. \end{aligned}$$

Sejam  $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})^\top$  e  $\tilde{\boldsymbol{\theta}} = (\mu_0, \tilde{\sigma})^\top$  os estimadores de máxima verossimilhança de  $\boldsymbol{\theta} = (\mu, \sigma)^\top$ , sendo  $\hat{\boldsymbol{\theta}}$  o estimador de máxima verossimilhança irrestrito e  $\tilde{\boldsymbol{\theta}}$  o estimador de máxima verossimilhança restrito sob a hipótese nula.

Nesse contexto, as formas das estatísticas robustas de Wald, escore de Rao e gradiente são dadas, respectivamente, por

$$S_W^* = n(\hat{\mu} - \mu_0)^2 \left[ \hat{\mathbf{H}}_n(\tilde{\boldsymbol{\theta}}) \hat{\mathbf{J}}_n(\tilde{\boldsymbol{\theta}})^{-1} \hat{\mathbf{H}}_n(\tilde{\boldsymbol{\theta}}) \right]_{\mu\mu},$$

$$S_R^* = \frac{1}{n\sigma^2} \left[ \sum_{l=1}^n m'(z_l) \right]^2 \left[ \hat{\mathbf{H}}_n(\tilde{\boldsymbol{\theta}}) \hat{\mathbf{J}}_n(\tilde{\boldsymbol{\theta}})^{-1} \hat{\mathbf{H}}_n(\tilde{\boldsymbol{\theta}}) \right]_{\mu\mu}$$

e

$$S_T^* = (\mu_0 - \hat{\mu}) \left[ \frac{1}{\sigma} \sum_{l=1}^n m'(z_l) \right] \left[ \hat{\mathbf{J}}_n(\tilde{\boldsymbol{\theta}})^{-1} \hat{\mathbf{H}}_n(\tilde{\boldsymbol{\theta}}) \right]_{\mu\mu}.$$

Para mais detalhes e exemplos associados a testes de hipóteses compostas, pode-se consultar Kent (1982) e Lemonte (2013).

# 5 Modelos de Regressão sob Má especificação

Nesta capítulo introduzimos o contexto de má especificação em modelos de regressão. Nesse contexto, apresentamos a forma geral para teste de hipótese (tanto para hipótese simples, quanto para composta) quando estamos interessados em testar um vetor de parâmetros associado a um modelo de regressão. Nosso objetivo, neste capítulo, é apresentar expressões para os testes de hipóteses dos parâmetros do modelo de regressão quando estamos na presença de má especificação.

## 5.1 Modelos de Regressão e Má especificação

Em geral, um modelo de regressão linear têm dois propósitos: realizar predição e/ou previsão e explicar a relação entre uma variável dependente (dita resposta e representada por  $Y$ ) e uma ou mais variáveis independentes (ditas covariáveis e representadas por  $X$ ) que se deseja estudar. Considere o modelo de regressão

$$Y_l = \beta_0 + \beta_1 x_l + Z_l, \quad l = 1, \dots, n,$$

sendo  $n$  o número de observações,  $\beta_0$  e  $\beta_1$  parâmetros desconhecidos a serem estimados, os erros denotados por  $Z_l$ , os quais assumimos independentes e normalmente distribuídos com média zero e variância  $\sigma^2$ . Consequentemente,

$$Y_l | x_l \sim N(\beta_0 + \beta_1 x_l, \sigma^2),$$

com  $Y_l$  independente de  $Y_j$ ,  $\forall l \neq j$ .

Os parâmetros do modelo podem ser estimados por máxima verossimilhança. Os estimadores apresentam boas propriedades e as técnicas usuais utilizadas em análise de regressão são bastante eficientes. Para mais detalhes sugerimos a leitura de Montgomery, Peck e Vining (2015).

No entanto, se a distribuição assumida para o erro (e, consequentemente, para a variável resposta) é mal especificada, as inferências para os parâmetros do modelo são fortemente influenciadas. Nesse contexto, dizemos que o modelo está mal especificado.

Considere que para cada indivíduo de uma amostra de tamanho  $n$  tenhamos um vetor de covariáveis  $\mathbf{x}_l = (x_{1l}, x_{2l}, \dots, x_{pl})^\top$ , em que  $x_{1l} = 1$ ,  $\forall l = 1, \dots, n$ , e uma variável resposta  $Y_l$ . Associado ao vetor de covariáveis  $\mathbf{x}_l = (x_{1l}, x_{2l}, \dots, x_{pl})^\top$ , temos o vetor de constantes conhecidas  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)^\top$ . Nesse contexto, o modelo de regressão

associado à variável resposta  $Y_l$  e às covariáveis  $\mathbf{x}_l$  é dado por

$$\begin{aligned} Y_l &= \gamma_1 x_{1l} + \gamma_2 x_{2l} + \dots + \gamma_p x_{pl} + \sigma_0 Z_l \\ &= \mathbf{x}_l^\top \boldsymbol{\gamma} + \sigma_0 Z_l, \end{aligned}$$

em que  $Z_1, \dots, Z_n$  são variáveis aleatórias independentes e identicamente distribuídas com função densidade conhecida. E, nesse contexto, temos que  $Y_l \mid \mathbf{x}_l$  são variáveis aleatórias independentes com distribuição  $g_{Y_l \mid \mathbf{x}_l}(y_l \mid \mathbf{x}_l)$ .

No entanto, iremos assumir que o modelo de regressão associado à variável resposta  $Y_l$  e as covariáveis  $\mathbf{x}_l$  é dado por

$$\begin{aligned} Y_l &= \beta_1 x_{1l} + \beta_2 x_{2l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l \\ &= \mathbf{x}_l^\top \boldsymbol{\theta} + \sigma_0 Z_l, \end{aligned}$$

em que  $\boldsymbol{\theta} = (\beta_1, \dots, \beta_p)^\top \in \Theta \subset \mathbb{R}^p$  é o vetor de parâmetros desconhecido associado ao vetor de covariáveis  $\mathbf{x}_l$ , e  $\sigma_0$  é conhecido. Assumimos que  $Z_1, \dots, Z_n$  são variáveis aleatórias independentes e identicamente distribuídas com função densidade  $f_0(\cdot)$  e que  $Y_l \mid \mathbf{x}_l$  são variáveis aleatórias independentes com densidade  $f_{Y_l \mid \mathbf{x}_l}(y_l \mid \mathbf{x}_l; \boldsymbol{\theta})$  dada por

$$f_{Y_l \mid \mathbf{x}_l}(y_l \mid \mathbf{x}_l; \boldsymbol{\theta}) = \frac{1}{\sigma_0} f_0 \left( \frac{y_l - \mathbf{x}_l^\top \boldsymbol{\theta}}{\sigma_0} \right).$$

Ou seja, consideramos que a distribuição de  $Y_l \mid \mathbf{x}_l$  pertence à família de posição e escala, descrita na Seção 2, com parâmetro de posição  $\mu(\mathbf{x}_l, \boldsymbol{\theta}) = \mathbf{x}_l^\top \boldsymbol{\theta}$ , e parâmetro de escala  $\sigma_0$  conhecido.

Assim, dada a amostra observada de tamanho  $n$ , a função de verossimilhança para o vetor de parâmetros desconhecido  $\boldsymbol{\theta}$  é dada por

$$L(\boldsymbol{\theta}) = \prod_{l=1}^n \frac{1}{\sigma_0} f_0 \left( \frac{y_l - \mathbf{x}_l^\top \boldsymbol{\theta}}{\sigma_0} \right),$$

em que  $y_l$  é o valor observado de  $Y_l$  e  $\sigma_0$  é o parâmetro de escala de  $Y_l$ , que estamos supondo conhecido.

Neste caso, fazendo  $z_l = \sigma_0^{-1}(y_l - \mathbf{x}_l^\top \boldsymbol{\theta})$  e  $m(z_l; \mathbf{x}_l, \boldsymbol{\theta}) = \log f_0(z_l)$ , então a função de log-verossimilhança é dada por

$$\ell_n(\boldsymbol{\theta}) = \sum_{l=1}^n m(z_l; \mathbf{x}_l, \boldsymbol{\theta}) - n \log \sigma_0. \quad (5.1)$$

O estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  é o vetor de parâmetros  $\hat{\boldsymbol{\theta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^\top$  que maximiza (5.1) em  $\Theta$  e converge para  $\boldsymbol{\theta}_g = (\beta_{g1}, \dots, \beta_{gp})^\top \in \Theta$ , o vetor de parâmetros que minimiza KLIC definido na Seção 3.2.1.

Defina

$$\begin{aligned}\mathbf{u}^{(l)}(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} m(z_l; \mathbf{x}_l, \boldsymbol{\theta}) = \left( \frac{\partial}{\partial \beta_j} m(z_l; \mathbf{x}_l, \boldsymbol{\theta}) \right)_{j=1, \dots, p} \\ &= \left( -\frac{1}{\sigma_0} \frac{\partial m(z_l; \mathbf{x}_l, \boldsymbol{\theta})}{\partial z_l} x_{jl} \right)_{j=1, \dots, p}, \quad l = 1, \dots, n,\end{aligned}$$

e, neste caso, vetor escore é dado por

$$\mathbf{u}_n(\boldsymbol{\theta}) = \sum_{l=1}^n \mathbf{u}^{(l)}(\boldsymbol{\theta}),$$

cujos componentes são

$$\mathbf{u}_{n\beta_j}(\boldsymbol{\theta}) = -\frac{1}{\sigma_0} \sum_{l=1}^n \frac{\partial m(z_l; \mathbf{x}_l, \boldsymbol{\theta})}{\partial z_l} x_{jl}.$$

Sendo assim, baseando-se na notação de Medeiros et al. (2014), podemos escrever

$$\mathbf{u}_n(\boldsymbol{\theta}) = \sigma_0^{-1} \mathbf{X}^\top \mathbf{a},$$

com  $\mathbf{a} = (a_1, a_2, \dots, a_n)^\top$ ,  $a_l = -m'(z_l)$ ,  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)^\top$  a matriz de covariáveis de dimensão  $n \times p$  e  $l = 1, \dots, n$  e  $m'(z_l) = \partial m(z_l; \mathbf{x}_l, \boldsymbol{\theta}) / \partial z_l$ .

Matricialmente, temos

$$\mathbf{u}^{(l)}(\boldsymbol{\theta}) = \begin{bmatrix} -\frac{1}{\sigma_0} m'(z_l) x_{1l} \\ \frac{1}{\sigma_0} m'(z_l) x_{2l} \\ \vdots \\ -\frac{1}{\sigma_0} m'(z_l) x_{pl} \end{bmatrix}.$$

Agora, consideremos as matrizes  $\mathbf{H}_1$  e  $\mathbf{H}_2$ , de dimensões  $n \times n$ , definidas por

$$\mathbf{H}_1 = \begin{bmatrix} [m'(z_1)]^2 & 0 & \cdots & 0 \\ 0 & [m'(z_2)]^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [m'(z_n)]^2 \end{bmatrix}$$

e

$$\mathbf{H}_2 = \begin{bmatrix} m''(z_1) & 0 & \cdots & 0 \\ 0 & m''(z_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m''(z_n) \end{bmatrix}.$$

Dessa forma, de (3.11) e (3.12), temos que

$$\hat{\mathbf{H}}_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{l=1}^n \frac{\partial \mathbf{u}^{(l)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = (-n\sigma_0^2)^{-1} \mathbf{X}^\top \mathbf{H}_2 \mathbf{X} \quad (5.2)$$

e

$$\hat{\mathbf{J}}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{l=1}^n \mathbf{u}^{(l)}(\boldsymbol{\theta}) \mathbf{u}^{(l)}(\boldsymbol{\theta})^\top = (n\sigma_0^2)^{-1} \mathbf{X}^\top \mathbf{H}_1 \mathbf{X}. \quad (5.3)$$

Note que os estimadores  $\hat{\mathbf{J}}_n(\boldsymbol{\theta})$  e  $\hat{\mathbf{H}}_n(\boldsymbol{\theta})$  estão em função das matrizes  $\mathbf{X}$ ,  $\mathbf{H}_1$  e  $\mathbf{H}_2$ .

## 5.2 Teste de Hipóteses

Suponha que o interesse seja testar  $\mathcal{H}_0: \boldsymbol{\theta}_g = \boldsymbol{\theta}_0$  versus  $\mathcal{H}_1: \boldsymbol{\theta}_g \neq \boldsymbol{\theta}_0$ , em que  $\boldsymbol{\theta}_0 = (\beta_{10}, \dots, \beta_{p0})^\top$  é um vetor de constantes. Neste caso, podemos utilizar as estatísticas

$$S_W = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{H}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

$$S_R = n^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \mathbf{J}(\hat{\boldsymbol{\theta}}_0)^{-1} \mathbf{u}_n(\boldsymbol{\theta}_0)$$

e

$$S_T = \mathbf{u}_n(\boldsymbol{\theta}_0)^\top \mathbf{J}(\hat{\boldsymbol{\theta}}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Sob a hipótese nula,  $S_W$ ,  $S_R$  e  $S_T$  tem distribuição aproximadamente qui-quadrado com  $p$  graus de liberdade, para  $n$  grande (ver Kent (1982) e Lemonte (2013)).

Agora, considere a partição  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  e, da mesma forma,  $\boldsymbol{\theta}_g = (\boldsymbol{\psi}_g^\top, \boldsymbol{\lambda}_g^\top)^\top$ , em que  $\boldsymbol{\psi}$  e  $\boldsymbol{\lambda}$  tem dimensão  $q$  e  $p - q$ , respectivamente. Neste caso, note que  $\boldsymbol{\psi}^\top = (\beta_1, \beta_2, \dots, \beta_q)^\top$  e  $\boldsymbol{\lambda}^\top = (\beta_{q+1}, \beta_{q+2}, \dots, \beta_p)^\top$ . Os parâmetros  $\boldsymbol{\psi}_g^\top = (\beta_{g1}, \beta_{g2}, \dots, \beta_{gq})^\top$  e  $\boldsymbol{\lambda}_g^\top = (\beta_{g(q+1)}, \beta_{g(q+2)}, \dots, \beta_{gp})^\top$  são conforme definidos na Seção 3.2.1.

Suponha, agora, que o interesse seja testar a hipótese nula  $\mathcal{H}_0: \boldsymbol{\psi}_g = \boldsymbol{\psi}_0$  contra a hipótese alternativa  $\mathcal{H}_1: \boldsymbol{\psi}_g \neq \boldsymbol{\psi}_0$ , em que  $\boldsymbol{\psi}_0$  é um valor especificado e  $\boldsymbol{\lambda}$  é o vetor de parâmetros de perturbação. A partição de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  induz as partições das matrizes  $\mathbf{H}$  e  $\mathbf{J}$  e do vetor  $\mathbf{u}_n(\boldsymbol{\theta})$ , que seguem conforme apresentado em (3.13), (3.14) e (3.15).

Da mesma forma, seja  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\lambda}}^\top)^\top$  e  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\psi}_0^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$  os estimadores de máxima verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\lambda}^\top)^\top$  irrestrito e restrito (obtido sob a hipótese nula), respectivamente. Estes estimadores satisfazem  $\mathbf{u}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$  e  $\mathbf{u}_{n\lambda}(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ .

As estatísticas Wald robusta, escore robusta e gradiente robusta para testar a hipótese  $\mathcal{H}_0: \boldsymbol{\psi}_g = \boldsymbol{\psi}_0$  são dadas, respectivamente, por

$$S_W = n(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)^\top [\mathbf{H}(\tilde{\boldsymbol{\theta}}) \mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\boldsymbol{\psi}\boldsymbol{\psi}} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

$$S_R = n^{-1} \mathbf{u}_{n\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}})^\top \mathbf{H}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}) [\mathbf{H}(\tilde{\boldsymbol{\theta}}) \mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\boldsymbol{\psi}\boldsymbol{\psi}} \mathbf{H}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}) \mathbf{u}_{n\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}})$$

e

$$S_T = \mathbf{u}_{n\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}})^\top [\mathbf{J}(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}})]_{\boldsymbol{\psi}\boldsymbol{\psi}} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0).$$

Sob a hipótese nula, o limite em distribuição das estatísticas  $S_W$ ,  $S_R$  e  $S_T$  é qui-quadrado com  $q$  graus de liberdade, em que  $q$  representa a dimensão do vetor de parâmetros que estamos testando.



Como, na prática, as quantidades  $\mathbf{H}(\boldsymbol{\theta})$  e  $\mathbf{J}(\boldsymbol{\theta})$  são desconhecidas, pode-se utilizar os seus respectivos estimadores empíricos  $\hat{\mathbf{H}}_n(\boldsymbol{\theta})$  e  $\hat{\mathbf{J}}_n(\boldsymbol{\theta})$ , dados por (5.2) e (5.3), respectivamente. A hipótese nula é rejeitada se o valor observado da estatística de teste excede o quantil  $100(1 - \gamma)\%$  da distribuição  $\chi_q^2$ .

**Exemplo 5.2.1.** Modelo logístico (hipótese simples).

Suponha que temos um conjunto de 115 observações independentes da variável aleatória  $Y$ , com

$$Y_l = \gamma_1 x_{1l} + \gamma_2 x_{2l} + \sigma_0 Z_l,$$

sendo  $Z_l$  variáveis aleatórias independentes e identicamente distribuídas, em que  $Z_l \sim t(2, 0, 1)$ ,  $x_{1l} = 1 \forall l = 1, \dots, 115$ . Por outro lado, iremos analisar os dados supondo  $Z_l \sim \text{logística}(0, 1)$  com  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , e  $\sigma_0$  conhecido.

Neste caso, nosso interesse é testar as hipóteses

$$\mathcal{H}_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{versus} \quad \mathcal{H}_1 : \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

**Exemplo 5.2.2.** Modelo logístico (hipótese composta).

Suponha, agora, que temos um conjunto de 115 observações independentes da variável aleatória  $Y$ , com

$$Y_l = \gamma_1 x_{1l} + \gamma_2 x_{2l} + \gamma_3 x_{3l} + \sigma_0 Z_l,$$

sendo  $Z_l$  variáveis aleatórias independentes e identicamente distribuídas, em que  $Z_l \sim N(0, 1)$ . No entanto, iremos analisar os dados supondo  $Z_l \sim \text{logística}(0, 1)$ , com  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \sigma_0 Z_l$ , e  $\sigma_0$  conhecido.

Neste caso, nosso interesse é testar  $\mathcal{H}_0: \beta_2 = 0$  versus  $\mathcal{H}_1: \beta_2 \neq 0$ .

**Exemplo 5.2.3.** Modelo t-Student não padronizado.

Suponha, agora, que temos um conjunto de 115 observações independentes da variável aleatória  $Y$ , com

$$Y_l = \gamma_1 x_{1l} + \gamma_2 x_{2l} + \gamma_3 x_{3l} + \gamma_0 Z_l,$$

sendo  $Z_l$  variáveis aleatórias independentes e identicamente distribuídas, em que  $Z_l \sim \text{logística}(0,1)$ , mas iremos analisar os dados supondo  $Z_l \sim t(2,0,1)$ , com  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \sigma_0 Z_l$ , e  $\sigma_0$  conhecido.

Neste caso, nosso interesse é testar  $\mathcal{H}_0: \beta_2 = 0$  versus  $\mathcal{H}_1: \beta_2 \neq 0$ .

## 6 Simulações

Neste capítulo apresentamos os principais resultados numéricos obtidos através de simulações Monte Carlo, cujo objetivo principal é estudar o desempenho das estatísticas robustas tanto na família de posição e escala, para variáveis aleatórias independentes e identicamente distribuídas, quanto em modelos de regressão. Todas as simulações foram feitas no *software* estatístico R (R Core Team, 2014). Foi considerado 20000 replicas Monte Carlo. As maximizações foram feitas utilizando a função `optim` do R. Consideramos dois casos nos estudos de simulação: o primeiro caso tem como objetivo avaliar o desempenho das estatísticas usuais e robustas quando o interesse é testar hipótese simples e, para isso, consideramos diferentes modelos da família de posição e escala; em seguida, apresentamos resultados numéricos relacionado aos modelos de regressão apresentados no Capítulo 5 e, neste caso, consideramos testes de hipóteses tanto para hipótese simples, quanto para hipóteses compostas.

### 6.1 Família de Posição e Escala

Neste primeiro cenário, nosso interesse é testar  $\mathcal{H}_0: \theta = \theta_0$  versus  $\mathcal{H}_1: \theta \neq \theta_0$ . Nosso interesse é calcular para cada teste a taxa de rejeição da hipótese nula, isto é, o percentual de vezes em que as estatísticas correspondentes excedem o quantil de uma distribuição  $\chi_1^2$  para um nível de significância  $\alpha$  de 10%, 5% e 1%. Essa taxa é analisada tanto sob o modelo correto, quanto sob má especificação. Nesse contexto, por exemplo, supondo que o verdadeiro modelo é normal, mas assumimos o modelo logístico, as simulações são feitas seguindo dois passos:

1. (Modelo correto) Geramos dados da distribuição logística e calculamos a taxa de rejeição da hipótese nula considerando, para as expressões das estatísticas, o modelo logístico;
2. (Modelo mal especificado) Geramos dados da distribuição normal e calculamos a taxa de rejeição da hipótese nula considerando, para as expressões das estatísticas, o modelo logístico.

Ou seja, consideramos as mesmas expressões para as estatísticas e geramos dois conjuntos de dados diferentes. Todas as entradas das tabelas são porcentagens.

Inicialmente, consideramos o caso apresentado no Exemplo 4.1.2. Nesse caso, geramos uma amostra aleatória de tamanho  $n = 115$  de uma distribuição normal com parâmetros  $\mu = 1$  e  $\sigma = 1$ , mas vamos assumir que esses dados são de uma distribuição logística com

parâmetros  $\theta$  e  $\sigma = 1.5$ , que pertence a família de posição e escala. Os resultados são apresentados na Tabela 1.

Tabela 1 – Taxa de rejeição para testar  $\mathcal{H}_0: \theta = 1$ , com  $n = 115$ . Considerando a distribuição correta  $N(1, 1)$  e assumindo modelo *logístico*  $(\theta, 1.5)$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$LR$	10.50	5.13	0.81	0.03	0.00	0.00
$W$	10.75	5.29	0.85	0.04	0.00	0.00
$R$	10.64	5.23	0.84	0.03	0.00	0.00
$T$	10.59	5.14	0.80	0.03	0.00	0.00
$S_W^*$	10.60	4.95	0.77	10.21	4.91	0.94
$S_R^*$	10.67	5.10	0.74	10.12	4.79	0.92
$S_T^*$	10.66	5.01	0.74	10.12	4.87	0.92

Fonte: elaborada pelo autor.

Em seguida, consideramos um conjunto de  $n = 115$  observações de uma distribuição logística com parâmetros  $\mu = 0$  e  $\sigma = 2$ , mas assumimos que os dados são de uma distribuição t-Student não padronizada (2.5) com parâmetros  $\theta$ ,  $\sigma = 1$  e  $\nu = 2$ . Neste caso, as expressões das estatísticas robustas são apresentadas no Exemplo 4.1.4 e os resultados numéricos são apresentados na Tabela 2.

Tabela 2 – Taxa de rejeição para testar  $\mathcal{H}_0: \theta = 0$  com  $n = 115$ . Considerando a distribuição correta *logística*  $(0, 2)$  e assumindo modelo *t*  $(2, \theta, 1)$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$LR$	9.88	5.20	1.11	37.95	30.16	16.98
$W$	9.92	5.24	1.15	38.27	30.78	17.88
$R$	9.98	5.36	1.25	38.42	30.88	18.19
$T$	9.80	5.14	1.08	37.77	29.78	16.39
$S_W^*$	9.41	4.70	0.84	6.49	2.50	0.24
$S_R^*$	9.86	5.09	1.07	10.03	5.09	0.97
$S_T^*$	9.62	4.96	0.93	8.18	3.55	0.41

Fonte: elaborada pelo autor.

Finalmente, a Tabela 3 apresenta os resultados numéricos para o caso em que temos um conjunto de  $n = 115$  observações de uma distribuição logística com parâmetros  $\mu = 1$  e  $\sigma = 3$ , mas assumimos que os dados pertencem a uma distribuição normal com parâmetros  $\theta$  e  $\sigma = 1$ . As expressões das estatísticas robustas são apresentadas no Exemplo 4.1.3.

Como esperado, as estatísticas usuais  $LR$ ,  $W$ ,  $R$  e  $T$  apresentam bons resultados em todos os casos estudados, quando o modelo especificado para o conjunto de dados é o modelo correto. No entanto, quando isso não acontece, ou seja, quando o modelo que

assumimos para os dados não é, de fato, o modelo gerador dos dados, o desempenho dessas estatísticas é fortemente influenciado. Além disso, como não sabemos qual a distribuição destas estatísticas quando o modelo é mal especificado, podemos observar uma taxa de rejeição muito abaixo do esperado, ou seja, distantes dos níveis nominais considerados, como ocorre nos resultados apresentados na Tabela 1; ou uma taxa de rejeição muito alta, conforme apresentado nas Tabelas 2 e 3.

Tabela 3 – Taxa de rejeição para testar  $\mathcal{H}_0: \theta = 1$  com  $n = 115$ . Considerando a distribuição correta *logística* (1, 3) e assumindo modelo  $N(\theta, 1)$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$S$	10.05	5.15	0.94	76.58	71.31	64.07
$S^*$	10.17	5.08	0.87	10.21	4.79	1.05

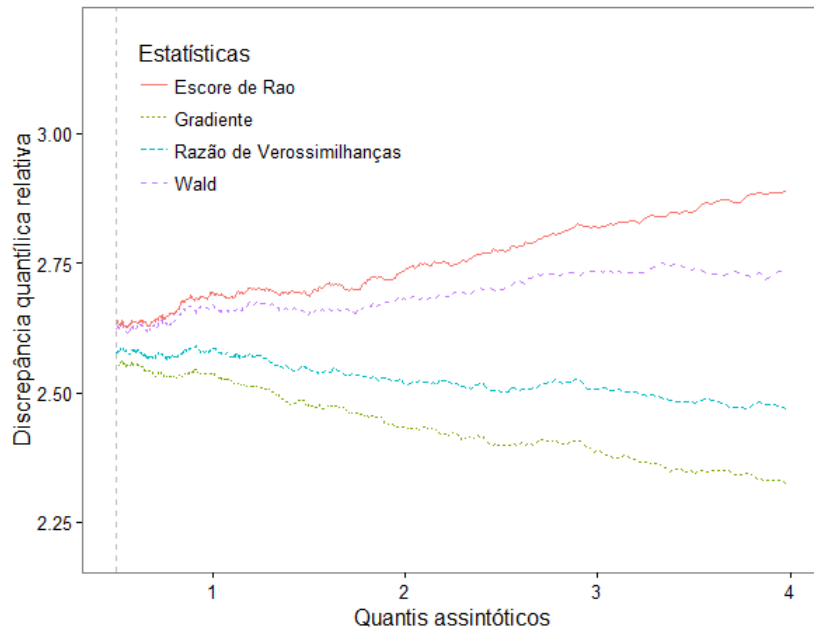
Fonte: elaborada pelo autor.

Por outro lado, as estatísticas robustas apresentaram bons resultados tanto quando o modelo especificado é correto, quanto na presença de má especificação. Isto é um indicativo de que as estatísticas, de fato, são robustas quanto à presença de má especificação e, quando o modelo especificado é correto, elas não sofrem perda de desempenho quando comparadas com as estatísticas usuais, o que é esperado. Vale notar, ainda, que, embora todas elas apresentem resultados satisfatórios, a versão robusta da estatística score de Rao tem um desempenho levemente superior, quando comparada com as versões robustas das estatísticas de Wald e gradiente.

Agora, consideramos o caso apresentado no Exemplo 4.1.4. Neste caso, geramos dados de uma distribuição logística com parâmetros  $\mu = 0$  e  $\sigma = 2$ , mas assumimos que os dados são de um modelo t-Student não padronizado com parâmetros  $\nu = 2$ ,  $\sigma = 1$  e  $\theta$ . O interesse é testar  $\mathcal{H}_0 : \theta = 0$ . A Figura 1 apresenta a curva da discrepância quantílica relativa *versus* o correspondente quantil assintótico das estatísticas  $S_W^*$ ,  $S_R^*$  e  $S_T^*$ . A discrepância quantílica relativa é definida como sendo a diferença entre o quantil exato (obtido por simulação) e o quantil assintótico, dividido pelo quantil assintótico. Quanto mais próximo de zero a discrepância quantílica relativa estiver, teremos maior indício que o limite em distribuição das estatísticas é qui-quadrado (para mais detalhes, ver Apêndice A).

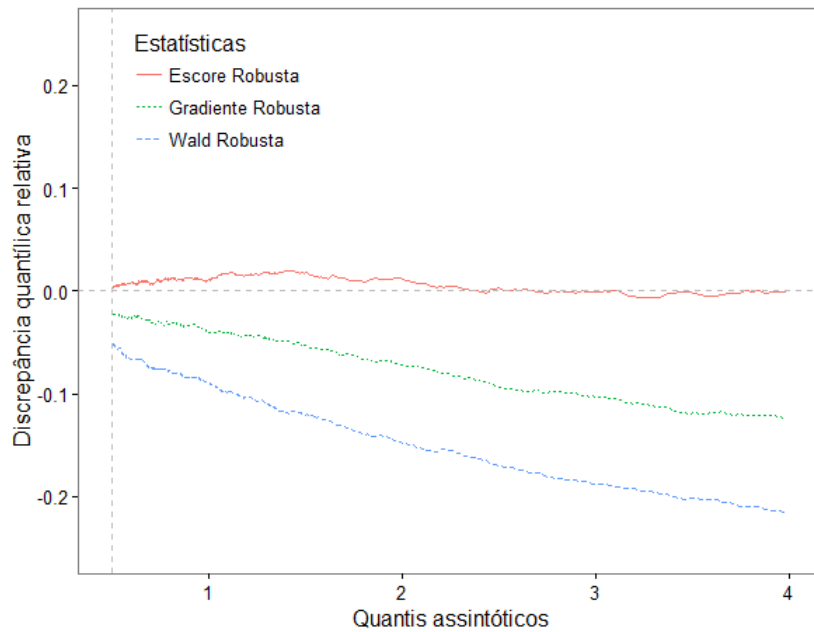
Com uma análise da Figura 1, notamos que a discrepância quantílica relativa das estatísticas usuais (LR, W, R e T) é relativamente distante do zero, indicando que a aproximação dessas estatísticas pela distribuição  $\chi_1^2$  pode não ser boa. De fato, quando erramos na especificação do modelo paramétrico para os dados, não sabemos a distribuição das estatísticas LR, W, R e T.

Figura 1 – Discrepância quantílica relativa para as estatísticas usuais testando  $\mathcal{H}_0: \theta = 0$ , considerando a distribuição *logística*  $(0, 2)$  e assumindo modelo  $t(2, \theta, 1)$ .



Fonte: elaborada pelo autor.

Figura 2 – Discrepância quantílica relativa para as estatísticas robustas testando  $\mathcal{H}_0: \theta = 0$ , considerando a distribuição *logística*  $(0, 2)$  e assumindo modelo  $t(2, \theta, 1)$ .



Fonte: elaborada pelo autor.

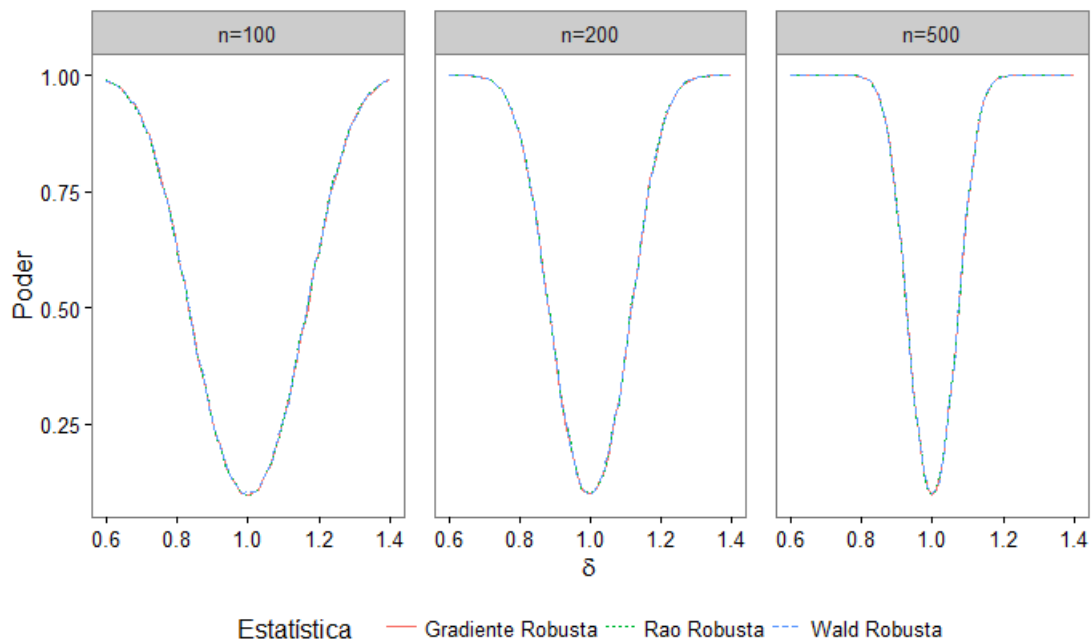
No entanto, a discrepância quantílica relativa das estatísticas robustas é bem próximo do zero, como podemos ver na Figura 2. Dessa forma, é razoável assumir que, em amostras

finitas, uma boa aproximação para a distribuição destas estatísticas seja qui-quadrado. Além disso, notamos que a versão robusta da estatística escore de Rao aparenta ser melhor aproximação para a distribuição qui-quadrado.

No que segue, apresentamos um breve estudo do poder dos testes robustos no caso em que temos observações de uma distribuição normal de parâmetros  $\mu = 1$  e  $\sigma^2 = 1$ , mas assumimos que os dados são de um modelo logístico de parâmetros  $\theta$  e  $\sigma_1^2 = 1$ , conforme apresentado no Exemplo 4.1.2. Neste caso, nosso interesse é testar  $\mathcal{H}_0 : \theta = 1$ . Consideramos  $n = 100, 200, 500$  e  $\alpha = 5\%$ . Para simular o poder do teste robusto, computamos a taxa de rejeição sob a hipótese alternativa  $\mathcal{H}_1 : \theta = \delta$  para diferentes valores de  $\delta$ .

Os gráficos de poder dos testes robustos considerando diferentes tamanhos de amostra são apresentados na Figura 3. Como podemos observar, não existe um teste uniformemente mais poderoso que o outro. Na verdade, as curvas de poder são, praticamente, indistinguíveis entre as estatísticas robustas.

Figura 3 – Poder do teste robusto para  $\alpha = 5\%$ , com tamanho de amostra  $n = 100, 200, 500$ . Considerando o modelo correto  $N(1, 1)$ , mas assumindo o modelo *logístico*  $(\theta, 1)$ .



Fonte: elaborada pelo autor.

## 6.2 Modelos de Regressão

No contexto de modelos de regressão, consideramos um estudo dos exemplos apresentados na Seção 5.2. As covariáveis utilizadas nas simulações são obtidas da distribuição normal ou uniforme.

Inicialmente, consideramos o caso apresentado no Exemplo 5.2.1, cujo modelo é  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , com  $Z_l \sim t(2, 0, 1)$ , mas vamos supor  $Z_l \sim \text{logística}(0, 1)$ , com  $\sigma_0$  conhecido. Neste caso, testamos a hipótese

$$\mathcal{H}_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

e computamos a taxa de rejeição (%), que pode ser vista na Tabela 4. Como esperado, notamos que as estatísticas usuais apresentam um bom desempenho quando o modelo especificado é correto, o que não acontece quando o modelo é mal especificado. Neste último caso notamos que essas estatísticas são visivelmente influenciadas pela má especificação do modelo. Por outro lado, as estatísticas robustas apresentam resultados relativamente bons, principalmente a versão robusta da estatística escore.

Tabela 4 – Taxa de rejeição para testar hipótese simples dada no Exemplo 5.2.1 com  $n = 115$  para o modelo  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , considerando  $Z_l \sim t(2, 0, 1)$  e assumindo  $Z_l \sim \text{logística}(0, 1)$ , com  $\sigma_0$  fixo e  $x_{1l} = 1, \forall l$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$LR$	10.24	5.01	1.11	1.77	0.61	0.02
$W$	10.32	5.12	1.16	1.80	0.64	0.03
$R$	10.47	5.24	1.23	1.87	0.66	0.03
$T$	10.13	4.95	1.09	1.75	0.58	0.02
$S_W^*$	8.83	4.22	0.63	9.17	4.32	0.66
$S_R^*$	9.75	4.88	0.84	9.68	4.66	0.83
$S_T^*$	9.31	4.52	0.74	9.40	4.44	0.75

Fonte: elaborada pelo autor.

No que segue, estamos interessados em testar hipóteses compostas, como é o caso do Exemplo 5.2.2, em que o erro tem distribuição normal padrão e assumimos logística padrão. Note, por meio da Tabela 5, que as estatísticas robustas apresentam um ótimo desempenho tanto no contexto de modelo mal especificado, quanto no contexto de modelo correto. E, conforme esperado, as estatísticas usuais são fortemente penalizadas pela má especificação do modelo.

Agora, consideramos o Exemplo 5.2.3, que temos um conjunto de 115 observações e um modelo de regressão dado por  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \sigma_0 Z_l$ , com  $Z_l \sim \text{logística}(0, 1)$ , mas que iremos analisar os dados supondo  $Z_l \sim t(2, 0, 1)$ , com  $\sigma_0$  conhecido. Neste caso, os resultados estão apresentados na Tabela 6.

Tabela 5 – Taxa de rejeição para testar hipótese  $\mathcal{H}_0: \beta_2 = 0$  com  $n = 115$  do modelo  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \sigma_0 Z_l$ , considerando  $Z_l \sim N(0, 1)$  e assumindo  $Z_l \sim \text{logística}(0, 1)$ , com  $\sigma_0$  fixo e  $x_{1l} = 1, \forall l$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$LR$	10.17	4.81	1.10	1.08	0.21	0.00
$W$	10.27	4.92	1.14	1.11	0.22	0.00
$R$	10.45	4.99	1.21	1.15	0.22	0.00
$T$	10.12	4.78	1.07	1.08	0.21	0.00
$S_W^*$	10.87	5.26	0.91	10.61	5.51	0.87
$S_R^*$	11.47	5.62	1.18	10.95	5.71	1.03
$S_T^*$	10.75	5.15	0.95	10.51	5.47	0.90

Fonte: elaborada pelo autor.

Tabela 6 – Taxa de rejeição para testar hipótese  $\mathcal{H}_0: \beta_2 = 0$  com  $n = 115$  do modelo  $Y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l} + \sigma_0 Z_l$ , considerando  $Z_l \sim \text{logística}(0, 1)$  e assumindo  $Z_l \sim t(2, 0, 1)$ , com  $\sigma_0$  fixo e  $x_{1l} = 1, \forall l$ .

Estatísticas	Modelo correto			Modelo mal especificado		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
$LR$	9.84	5.21	1.09	17.62	11.17	3.76
$W$	11.89	6.35	1.60	24.62	16.56	7.08
$R$	11.16	6.00	1.49	20.21	13.06	5.16
$T$	10.63	5.44	1.14	20.57	13.05	4.38
$S_W^*$	11.00	5.54	0.75	12.62	5.94	0.97
$S_R^*$	12.49	6.84	1.54	13.44	7.07	1.76
$S_T^*$	10.20	5.24	0.71	10.55	5.01	0.73

Fonte: elaborada pelo autor.

Observe, com base na Tabela 6, que as estatísticas robustas apresentam bons resultados, comparando-se com as estatísticas usuais, no contexto de má especificação. Quando o modelo especificado é correto, as estatísticas usuais, tal como as estatísticas robustas, apresentam, em geral, bom desempenho. Além disso, neste caso, a estatística gradiente robusta apresenta uma melhor performance que as demais.

Por fim, consideramos o modelo  $Y_l = \beta_1 x_{1l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l$  com  $Z_l \sim N(0, 1)$ , mas assumimos  $Z_l \sim \text{logística}(0, 1)$  e estamos interessados em testar  $\mathcal{H}_0: \beta_p = 0$ , para diferentes valores de  $p$  e considerando diferentes tamanho de amostra ( $n = 115, 200, 500$ ). Nosso objetivo é avaliar o desempenho destas estatísticas quando aumentamos o número de parâmetros no modelo. Nesse contexto, computamos a taxa de rejeição (%) da hipótese nula para um nível de significância  $\alpha$  de 10%, 5% e 1%, apenas sob modelo mal especificado. Os resultados estão apresentados na Tabela 7.

Como podemos observar na Tabela 7, conforme aumentamos o número de parâme-



Tabela 7 – Efeito do aumento do número de parâmetros nas estatísticas robustas para testar  $\mathcal{H}_0: \beta_p = 0$  considerando o modelo  $Y_l = \beta_1 x_{1l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l$  com  $l = 1, \dots, n$ ,  $Z_l \sim N(0, 1)$ ,  $n = 115, 200, 500$ , mas assumindo  $Z_l \sim \text{logística}(0, 1)$  e  $x_{1l} = 1, \forall l$ .

$n$	$p$	$\alpha = 1\%$			$\alpha = 5\%$			$\alpha = 10\%$		
		$S_W^*$	$S_R^*$	$S_T^*$	$S_W^*$	$S_R^*$	$S_T^*$	$S_W^*$	$S_R^*$	$S_T^*$
115	2	0.77	0.93	0.82	5.04	5.25	5.12	9.77	10.05	9.76
	3	0.97	1.07	1.00	5.47	5.66	5.45	10.59	10.91	10.62
	4	1.34	1.60	1.06	5.86	6.12	5.59	11.31	11.75	11.03
	5	1.18	1.35	1.20	6.57	6.90	5.79	12.10	12.38	11.30
	6	1.57	1.71	1.39	6.61	6.95	6.25	13.20	13.47	12.22
	7	2.21	2.49	1.23	7.37	7.59	6.78	12.72	12.98	12.06
	8	1.90	2.08	1.52	8.07	8.30	6.95	13.24	13.53	12.04
	200	2	0.88	0.99	0.91	5.07	5.30	5.11	9.78	9.93
3		0.92	0.99	0.94	5.38	5.50	5.39	10.44	10.55	10.44
4		0.94	1.04	0.94	5.16	5.38	5.10	10.96	11.10	10.78
5		1.05	1.18	0.99	5.75	5.93	5.59	11.30	11.48	10.69
6		1.35	1.44	1.22	6.38	6.50	5.97	10.90	11.04	10.79
7		1.39	1.46	1.20	6.28	6.38	5.91	11.98	12.15	11.47
8		1.24	1.34	1.16	6.08	6.24	5.79	12.18	12.37	11.54
500		2	0.87	0.91	0.88	4.78	4.79	4.75	9.71	9.76
	3	0.96	1.01	0.96	4.91	4.96	4.90	10.12	10.15	10.07
	4	1.10	1.18	1.11	5.16	5.21	5.16	10.10	10.16	10.05
	5	1.01	1.03	1.00	5.46	5.49	5.39	10.19	10.24	10.10
	6	1.10	1.16	1.07	5.26	5.33	5.26	10.90	10.96	10.50
	7	1.14	1.17	1.10	5.34	5.42	5.12	11.25	11.31	10.80
	8	1.38	1.40	1.35	5.58	5.67	5.48	10.81	10.84	10.61

Fonte: elaborada pelo autor.

tros no modelo, em geral, a taxa de rejeição da hipótese nula aumenta, embora não aumente drasticamente. Além disso, percebemos que a estatística gradiente robusta aparenta uma taxa de rejeição menor que as demais, e isso se mantém à medida que aumentamos o número de parâmetros. Por exemplo, para o nível de significância de 10% e para  $n = 115$ , a taxa de rejeição da hipótese nula da estatística gradiente robusta aumenta de 9.97 % para 12.04%, quando aumentamos o número de parâmetros de 2 para 8.

Adicionalmente, notamos que, à medida que o tamanho da amostra aumenta de 115 para 500, a taxa de rejeição da hipótese nula é bem próxima dos níveis de significância considerados. De fato, este comportamento já era esperado, já que a medida que aumentamos o tamanho da amostra, melhoramos a aproximação da distribuição das estatísticas pela distribuição qui-quadrado.

Os resultados numéricos considerando as estatísticas usuais podem ser encontrados no Apêndice B e, como esperado, o desempenho dessas estatísticas é extremamente

prejudicado pelo efeito de má especificação.

Em geral, percebemos que a utilização das estatísticas robustas apresentam uma ótima performance no tocante a realização de testes de hipóteses na presença de má especificação. Estas estatísticas apresentaram bons resultados tanto no contexto de amostra aleatória independente e identicamente distribuída quanto no contexto de regressão.

## 7 Considerações Finais

Neste trabalho estudamos o que ocorre em termos inferenciais e, em particular, em testes de hipóteses, quando erramos na especificação do modelo, ou seja, quando o modelo que analisamos os dados não é, de fato, o modelo gerador dos dados. Apresentamos uma revisão bibliográfica dos principais estudos realizados no contexto de má especificação, mencionamos as principais referências nesta área e introduzimos estes conceitos em modelos de uma importante família de distribuições, a família de posição e escala.

Usando os resultados de Kent (1982) e Lemonte (2013), apresentamos as expressões para testes de hipóteses no contexto de má especificação para o caso de amostras independentes e identicamente distribuídas de membros da família de posição e escala e em modelos de regressão. Realizamos um estudo de simulação considerando essas expressões. Como esperado, as estatísticas robustas apresentaram um ótimo desempenho para testar hipóteses no contexto de má especificação, quando comparadas com as estatísticas usuais. Estas últimas são fortemente influenciadas pela presença de má especificação, apresentando taxas de rejeição muito acima (ou abaixo) do esperado. Além disso, as estatísticas robustas também são adequadas comparando-se com as estatísticas usuais, quando o modelo é correto. Também, vimos que nenhum destes testes (considerando as estatísticas robustas) é uniformemente mais poderoso que o outro e que, na verdade, em termos de poder, as estatísticas robustas são, praticamente, indistinguíveis.

Nestes estudos, percebemos que, em geral, a versão robusta da estatística score de Rao apresenta um desempenho levemente melhor que as demais, tanto na presença quanto na ausência de má especificação no contexto de amostra independente e identicamente distribuída. Por outro lado, quando estamos com uma estrutura de regressão, a versão robusta da estatística gradiente apresenta destaque, principalmente quando aumentamos o número de covariáveis no modelo. Contudo, embora tenhamos considerados diferentes modelos neste trabalho, entendemos que muitos outros estudos devem ser feitos para que se possa fazer alguma recomendação de uma dessas estatísticas (Wald robusta, score robusta e gradiente robusta) para testar hipóteses.

Percebemos que as estatísticas robustas continuam apresentando um bom desempenho quando o interesse é realizar testes de hipóteses para os parâmetros do modelo de regressão, considerando tanto hipóteses simples quanto compostas.

Por fim, concluímos que o uso das estatísticas robustas é eficaz no sentido de corrigir as consequências causadas pela má especificação do modelo. Na prática, nunca sabemos qual a verdadeira distribuição dos dados e, nesse sentido, não sabemos quando estamos errando na especificação do modelo. Com base nos resultados obtidos neste

trabalho, a sugestão, em termos de aplicação, é utilizar, para procedimentos inferenciais, tanto as estatísticas usuais, quanto as robustas e, no caso de conclusões divergentes, que pode indicar má especificação do modelo, notamos que as estatísticas robustas são mais confiáveis que as estatísticas usuais.

Uma possível continuação do que foi feito nesse trabalho é estudar de forma mais detalhada o comportamento destas estatísticas robustas no contexto de regressão, considerando diferentes modelos (e não só àqueles pertencentes a família de posição e escala) e com parâmetro de escala desconhecido. Além disso, um possível trabalho futuro pode considerar o uso destas estatísticas no contexto de dados de sobrevivência, considerando amostra censurada.

# APÊNDICE A – Gráfico da Discrepância Quantílica Relativa

Seja  $Y$  uma variável aleatória com função de distribuição acumulada  $F_Y(y) = Pr\{Y \leq y\}$ . O  $100\alpha\%$  percentil é dado pelo valor  $q$  tal que

$$F_Y(q) \geq \alpha \quad \text{e} \quad Pr(Y \geq q) \geq 1 - \alpha.$$

O gráfico da discrepância quantílica relativa, apresentado na Seção 6.1, é uma forma de avaliar se a distribuição das estatísticas são, sob a hipótese nula, aproximadamente  $\chi^2_{(\cdot)}$ .

O processo para construção do gráfico consiste em comparar os percentis exatos, que são obtidos via simulação (que serão denotados por  $v_\alpha$ ), e os percentis assintóticos, obtidos a partir da distribuição qui-quadrado (denotados por  $w_\alpha$ ). Desta forma, a discrepância quantílica relativa ( $DQR_\alpha$ ) é definida por

$$DQR_\alpha = \frac{v_\alpha - w_\alpha}{w_\alpha}.$$

Note que, se a diferença ( $v_\alpha - w_\alpha$ ) é pequena, temos que os percentis de nível  $\alpha$  são próximos, indicando que as distribuições comparadas são próximas. Se, por outro lado, essa distância aumenta, temos indícios de que as distribuições são distintas.

# APÊNDICE B – Efeito do aumento no número de parâmetros - Estatísticas Usuais

Tabela 8 – Efeito do aumento do número de parâmetros nas estatísticas usuais para testar  $\mathcal{H}_0 : \beta_p = 0$  considerando o modelo  $Y_l = \beta_1 x_{1l} + \dots + \beta_p x_{pl} + \sigma_0 Z_l$  com  $l = 1, \dots, n$ ,  $Z_l \sim N(0, 1)$ ,  $n = 115, 200, 500$ , mas assumindo  $Z_l \sim \text{logística}(0, 1)$  e  $x_{1l} = 1, \forall l$ .

$n$	$p$	$\alpha = 1\%$				$\alpha = 5\%$				$\alpha = 10\%$			
		$LR$	$W$	$R$	$T$	$LR$	$W$	$R$	$T$	$LR$	$W$	$R$	$T$
115	2	0.01	0.01	0.01	0.01	0.22	0.22	0.22	0.22	1.08	1.09	1.15	1.04
	3	0.00	0.02	0.02	0.02	0.00	0.21	0.28	0.23	0.00	0.94	1.11	1.02
	4	0.00	0.00	0.00	0.00	0.00	0.02	0.35	0.32	0.00	0.08	1.10	1.02
	5	0.00	0.00	0.00	0.00	0.00	0.02	0.23	0.21	0.00	0.18	1.24	1.20
	6	0.00	0.00	0.01	0.00	0.00	0.00	0.22	0.20	0.00	0.10	1.21	1.09
	7	0.00	0.00	0.00	0.00	0.00	0.04	0.29	0.23	0.00	0.10	1.31	1.22
	8	0.00	0.00	0.01	0.01	0.00	0.02	0.26	0.24	0.00	0.17	1.23	1.14
	200	2	0.02	0.02	0.02	0.02	0.27	0.28	0.29	0.26	1.14	1.16	1.18
3		0.00	0.00	0.00	0.00	0.00	0.22	0.27	0.26	0.00	1.00	1.06	1.03
4		0.00	0.00	0.01	0.00	0.00	0.01	0.22	0.20	0.00	0.17	1.08	1.05
5		0.00	0.00	0.01	0.00	0.00	0.02	0.26	0.23	0.00	0.22	1.18	1.16
6		0.00	0.00	0.01	0.01	0.00	0.04	0.27	0.26	0.00	0.16	1.22	1.19
7		0.00	0.00	0.00	0.00	0.00	0.02	0.29	0.26	0.00	0.14	1.23	1.18
8		0.00	0.00	0.00	0.00	0.00	0.01	0.32	0.30	0.00	0.10	1.16	1.09
500		2	0.02	0.02	0.02	0.02	0.27	0.28	0.28	0.26	1.07	1.08	1.08
	3	0.00	0.00	0.00	0.00	0.00	0.24	0.26	0.26	0.00	1.24	1.25	1.24
	4	0.00	0.00	0.01	0.01	0.00	0.02	0.28	0.26	0.00	0.14	1.03	1.01
	5	0.00	0.00	0.01	0.00	0.00	0.04	0.34	0.33	0.00	0.20	1.19	1.17
	6	0.00	0.00	0.00	0.00	0.00	0.03	0.30	0.30	0.00	0.16	1.01	1.00
	7	0.00	0.00	0.00	0.00	0.00	0.01	0.24	0.24	0.00	0.10	0.99	0.98
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.24	0.00	0.12	1.26	1.25

Fonte: elaborada pelo autor.

# Referências

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*. [S.l.: s.n.], 1973. p. 267–281.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. [S.l.]: SBM, 2001. v. 2.
- CASELLA, G.; BERGER, R. L. Inferência estatística-tradução da 2ª edição norteamericana. *Centage Learning*, 2011.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: *ABE-Projeto Fisher*. [S.l.]: Edgard Blücher, 2006.
- CORDEIRO, G. M. *Introdução à teoria de verossimilhança*. [S.l.]: Associação Brasileira de Estatística, 1992.
- CORDEIRO, G. M. *Introdução à teoria assintótica*. [S.l.]: Colóquio Brasileiro de Matemática, 1999.
- FRASER, D. et al. On information in statistics. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 36, n. 3, p. 890–896, 1965.
- KENT, J. T. Robust properties of likelihood ratio tests. *Biometrika*, Biometrika Trust, v. 69, n. 1, p. 19–27, 1982.
- KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The annals of mathematical statistics*, v. 22, n. 1, p. 79–86, 1951.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2011. v. 362.
- LEMONTE, A. J. On the gradient statistic under model misspecification. *Statistics & Probability Letters*, v. 83, n. 1, p. 390–398, 2013.
- LEMONTE, A. J. *The Gradient Test: Another Likelihood-Based Test*. [S.l.]: London: Academic Press, 2016.
- MEDEIROS, F. M. et al. Testing inference in accelerated failure time models. *International Journal of Statistics and Probability*, v. 3, n. 2, p. 121, 2014.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2015.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org/>>.
- RAO, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In: CAMBRIDGE UNIV PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.l.], 1948. v. 44, n. 01, p. 50–57.

- ROYALL, R.; TSOU, T.-S. Interpreting statistical evidence by using imperfect models: robust adjusted likelihood functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 65, n. 2, p. 391–404, 2003.
- SIINO, M.; FASOLA, S.; MUGGEO, V. M. Inferential tools in penalized logistic regression for small and sparse data: A comparative study. *Statistical Methods in Medical Research*, SAGE Publications, p. 0962280216661213, 2016.
- TERRELL, G. R. The gradient statistic. *Computing Science and Statistics*, v. 34, n. 34, p. 206–215, 2002.
- WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, v. 54, n. 3, p. 426–482, 1943.
- WHITE, H. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, p. 1–25, 1982.
- WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, v. 9, n. 1, p. 60–62, 1938.